THE SIGNIFICANCE OF HOST FACTORS IN MYCOBACTERIOPHAGE LYSOGENY

by

Curtis Carl Wadsworth

B.S. Bethany College, Bethany, WV, 1996

Submitted to the Graduate Faculty of

Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2002

UNIVERSITY OF PITTSBURGH

FACULTY OF ARTS AND SCIENCES

This dissertation was presented

by

Curtis Carl Wadsworth

It was defended on

22 November 2002

and approved by

Dr. Roger W. Hendrix

Dr. Linda Jen-Jacobson

Dr. James Pipas

Dr. Saleem Khan

Dr. Graham F. Hatfull
Dissertation Director

THE SIGNIFICANCE OF HOST FACTORS IN MYCOBACTERIOPHAGE LYSOGENY

Curtis Carl Wadsworth, PhD

University of Pittsburgh, 2002

Lysogeny is a defining feature of temperate bacteriophages. Temperate bacteriophages are able to establish lysogeny by integrating into the host chromosome or extrachromosomally, as a plasmid-like prophage in the host cytoplasm. In either case, host factors are involved in both the establishment and maintenance of lysogeny. Mycobacteriophage L5 forms an integrated prophage through the action of a phage-encoded tyrosine recombinase. L5 integrase (Int) binds to the phage attachment site, *attP*, and the bacterial attachment site, *attB*. Int binds *attP* bivalently by binding to the core, where strand exchange occurs, and to arm-type binding sites that flank the core. A host-encoded DNA binding protein, mIHF, is required for recombination and binds between the core and arm-type binding sites of *attP*. The bending of *attP* required for bivalent binding between the core and arm-type binding sites is thought to catalyzed by mIHF. The core of *attP* consists of a seven base pair overlap region flanked on either side by imperfect inverted repeats that make up the recombinase binding elements or RBEs. The RBEs were shown to be essential for core binding by creating *attP*s with mutations in the RBEs. When both of the RBEs are mutated Int can neither perform recombination, nor create specific complexes that involve core binding. When the right side RBE is wild type with mutant left side RBE, recombination can occur, and complexes involving core binding are observed. However, Int is unable to catalyze recombination when the left side RBE is wild type and the right is mutated, indicating

that contact must be made with *attP* on the right side of core for a stable Int/*attP* core complex to be formed. The Int protein consists of three domains. The C-terminal domain contains the catalytic residues and binds to *attP* and *attB* core, and the N-terminal domain binds the arm-type binding sites of *attP*. The middle domain connects the two outer domains, but its function in recombination is not well understood. To determine the function of the middle domain, mutations were made at conserved residues in the middle domain of Int, and these mutants were characterized. These mutants are able to catalyze recombination, but do not form the recombinagenic complexes observed by wild type Int. Increasing the mIHF concentration in these reactions augments the recombination efficiency and stabilizes recombinagenic complexes that involve Int/core binding. A wild phage isolated from a soil sample, CJW1, forms lysogens at 42°C but not 37°C. This phenotype could be due to a temperature sensitive repressor protein, however interactions with host factors that are only produced at 42°C could be necessary for the establishment of lysogeny.

FOREWORD

In my years at the University of Pittsburgh, I have had the opportunity to work and play with some of the greatest people in the world. I first walked into the Hatfull lab as a Hughes Fellow in the summer of 1995, and since that time, the enthuiasm for science and exuberance for life that seems to eminate from the lab and the people therein has made almost every day a pleasure. Graham Hatfull has been a great role model and mentor. Lori Bibb and John Lewis contributed to almost every aspect of my work for the last 5 years. I couldn't ask for better people to work with. Gary Sarkis and Carol Peña have been great role models and friends. Kathy Brendza and Andy Mackey helped me through some very trying times, and without their support I would probably be a garbage man right now. I could spend the next 5 pages listing the people that have helped me achieve this goal with their friendship and support, and that makes me very happy. Thank You All.

DEDICATION

This work is dedicated to my grandfather, Joe Dzopko. I will always try to live my life with the same devotion.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

I. INTRODUCTION

I.A. Bacteriophages

I.A.i. The Study of Bacteriophages

Bacteriophages make up the largest group of organisms in the biosphere. With an estimated 10^{31} bacteriophages (Hendrix et al. 1999), an appreciation of the diversity and complexity of these organisms can only be gauged by the current body of knowledge produced by their study. Relatively few phages have been studied in great detail, and yet, as can be noted in Appendix A, the study of bacteriophage has greatly enhanced our knowledge of microbiology, molecular biology and genetics. As new phages are discovered, our understanding of molecular biology and microbial genetics will continue to grow and new technologies will be developed that will continue to help molecular, cellular and developmental biologists understand the complexity of life.

I.B. Lysogeny

I.B.i. Lytic and Lysogenic Bacteriophage Lifstyles

Bacteriophages are obligate parasites of bacteria. Infection occurs when the bacteriophage (phage) binds to a specific receptor on the cell wall of the host cell with proteins at the end of its tail. This is a two step process involving reversible adsorbtion and irreversible injection (Bode and Kaiser 1965). Once the phage DNA has reached the cytoplasm, the phage DNA circularizes using short, complimentary, single-stranded regions of DNA at the 5′ end of the phage genome, called cohesive ends or cos sites,

forming nicked DNA circles referred to as Hershey circles, after A. D. Hershey who first characterized them for phages (Burgi et al. 1966). After circularization, the phage DNA is ligated to produce a circular episome (Fitzmaurice et al. 1984). Phages that are unable to circularize cannot replicate and therefore, cannot reproduce (Miller and Feiss 1988, Oyaski and Hatfull 1992, Xu and Feiss 1991).

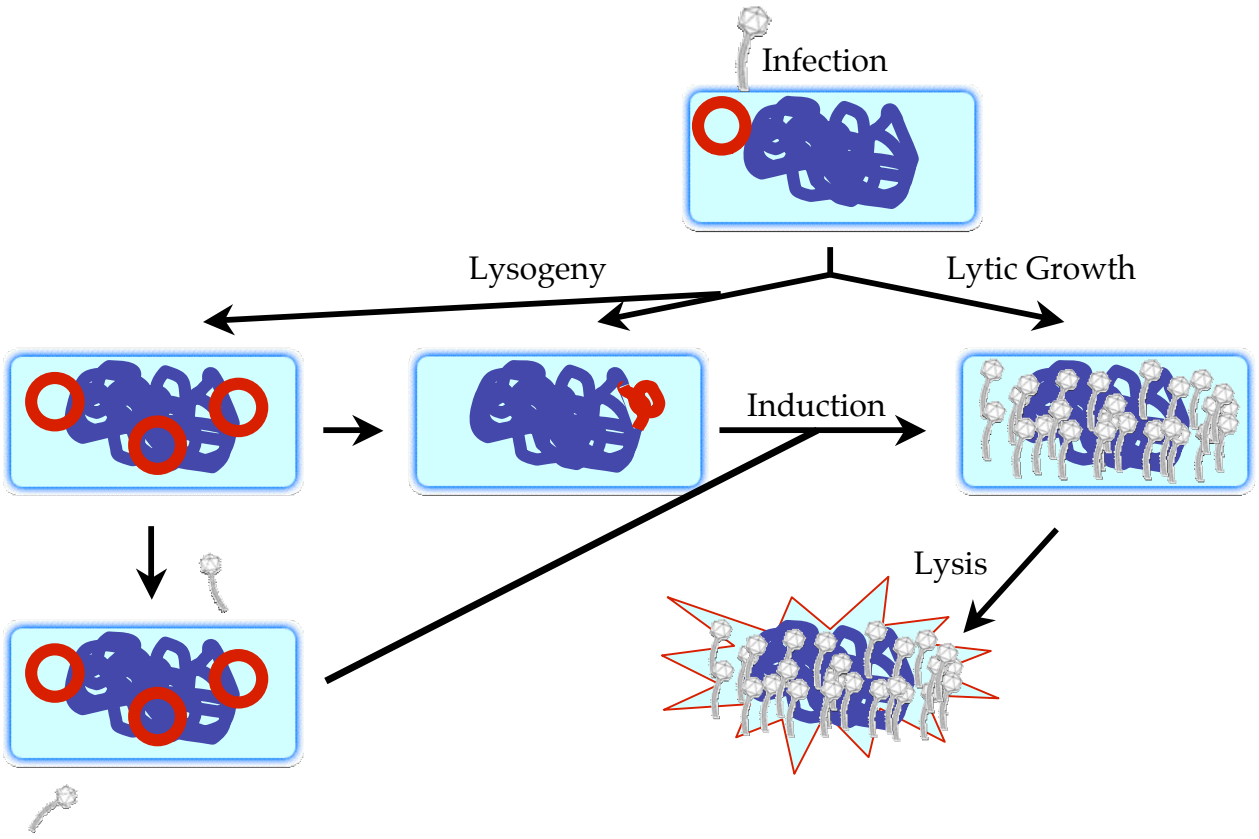
Upon entry into the cell and circularization of the phage genome, the phage has a "choice" of lifestyles. Lytic bacteriophages use the host's resources to perform rounds of phage replication, filling the bacterial cell with progeny phages before its inducing lysis, thereby releasing progeny phages into the environment. Temperate bacteriophages infect bacterial cells in the same way as lytic bacteriophages. However unlike lytic phages, temperate phages can undergo lysogeny, the process though which the phage genome can be maintained stably within the host cell as a prophage. A prophage can be maintained within an infected bacterial cell for many generations as part of the host chromosome, through integration, or as a plasmid in the cytoplasm, an extrachromosomal prophage. The prophage is replicated with the host DNA using the host's replication machinery.

A prophage may revert to vegetative growth at any time. Most lysogenic phages induce lysis of the infected cell in response to environmental factors, such as ultraviolet radiation and the presence of mutagens. Once induced, the phage's lytic genes are turned on, and lysogenic genes are repressed (Ptashne 1987, Ptashne et al. 1982). The phage takes over the biochemical machinery of the host cell to produce progeny phages. The host cell is then destroyed, releasing the progeny phages (Figure 1).

The processes of lysogenization and the maintenance of lysogeny depend on the presence of a phage-encoded repressor protein. The repressor binds to phage DNA and turns off lytic genes while allowing the expression genes required lysogeny

Figure 1. Lysogeny versus Lytic Growth

Bacteriophages infect their bacterial hosts by binding to a receptor on the host cell surface and injecting their compliment of DNA into the cytoplasm of the cell. Lytic phages take over the host cell's machinery following infection and begin to produce large numbers of progeny phage as shown in the panel on the far right. The phage then causes host cell lysis, releasing the progeny phage into the environment. The temperate lifestyle is shown on the left, and can be maintained least two ways. The prophage can be maintained as a low copy number plasmid in the cytoplasm of the host cell as shown on the far left. Lysogeny can also occur when the phage genome is inserted into the host chromosome, middle panels. The integrated prophage can remain part of the host chromosome for many generations and is replicated with the host chromosome. Some extrachormosomal lysogens can also perform integration under the right circumstances. Lysogeny can be reversed at any point during infection, producing a vegetative phage that undergoes lytic growth.


Infection
Lysogeny
Lytic Growth
Induction
Lysis

(Maniatis and Ptashne 1973, Ptashne 1973, Ptashne and Hopkins 1968). Lysogenic bacteria allow expression of the phage repressor at high enough levels to ensure that all phage genes are repressed, and production of repressor depends on the state of the host cell, and in many cases, host-encoded proteases degrade the repressor, as is the case in lambda where host encoded RecA binds to and cleaves λ repressor (Craig and Roberts 1980, Phizicky and Roberts 1980, Roberts and Roberts 1975, Roberts et al. 1978). If the concentration of the repressor drops below a certain threshold, induction occurs.

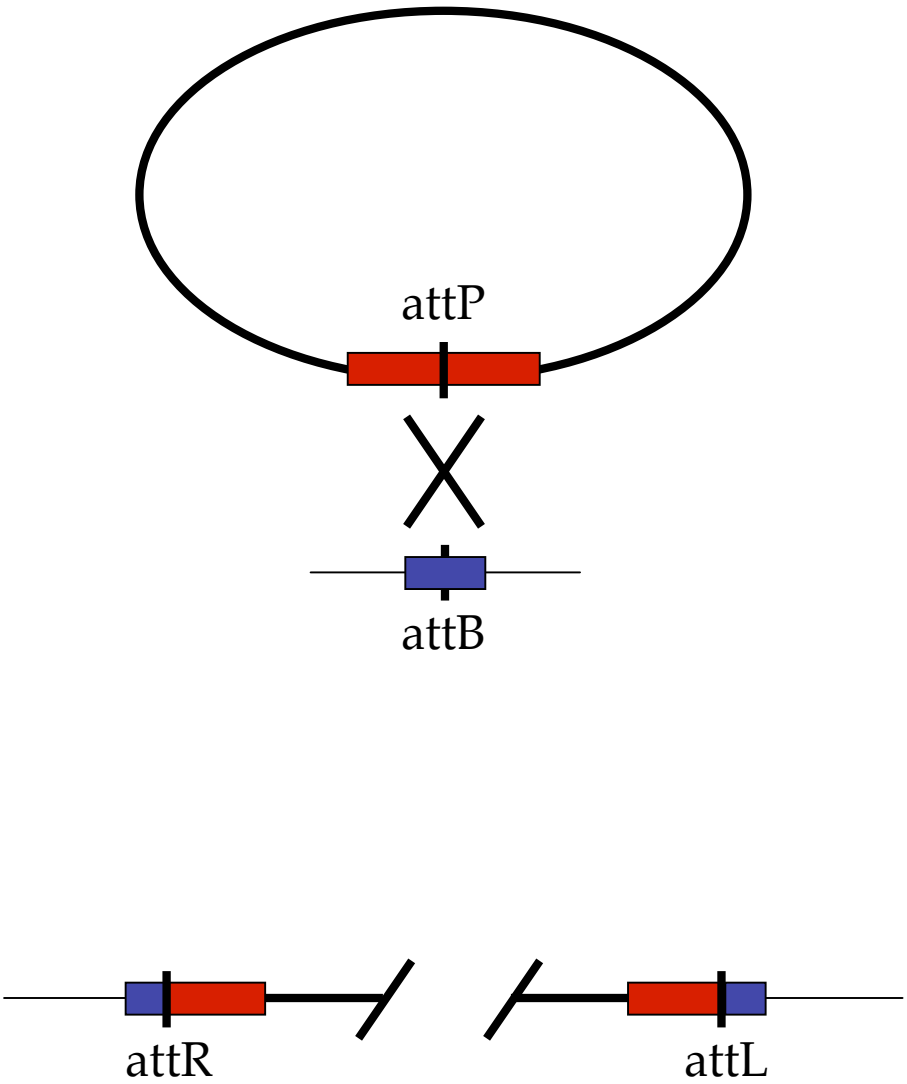
I.B.ii. Integration

Integrative recombination is the process by which the bacteriophage genome is inserted into the bacterial chromosome. Phage integration is usually site-specific using phage encoded tyrosine recombinases (Figure 2) (Nash 1975, Nash and Robertson 1981), although Mu has developed an integration system that can place a prophage in almost any bacterial gene using transposition (Mizuuchi and Mizuuchi 1993, Symonds et al. 1987), and filamentous phage CTX appears to use host encoded enzymes to perform site-specific recombination (Huber and Waldor 2002).

Site-specific recombination occurs between the phage attachment site, *attP*, in the phage genome and the bacterial attachment site, *attB*, found on the host chromosome. Tyrosine recombinases utilize an active site tyrosine to perform nucleophilic attack on the DNA substrates, but an active site serine can also be used to perform the nucleophilic attack in resolvase-like integrases as in phage phiC31 (Thorpe and Smith 1998, Thorpe et al. 2000). Host-encoded factors are sometimes utilized in conjunction with integrases to create elaborate protein/DNA complexes that are necessary for recombination.

Figure 2. Site-Specific Integration of the Phage Genome

Once the phage DNA has circularized, a phage encoded integrase binds the phage DNA at the phage attachment site, *attP*, and the host chromosome at the bacterial attachment site, *attB*. *attP* and *attB* are homologous at the core binding region and at a 6-8 basepair overlap region within the core where strand exchange will occur. Host factors are sometimes part of the integrase/*attP*/*attB* complex and help to form complicated protein/DNA structures. When the proper protein/DNA complex has been built, integrase catalyzes a site-specific recombination event that inserts the phage DNA into the host chromosome. This process is reversed during excision when the phage-encoded excisionase is produced.


attP
attB
attR
attL

I.B.iii. Extrachromosomal Prophages

Prophages of some temperate bacteriophages are maintained extrachromosomally. Extrachromosomal prophages are plasmid-like episomes and have been observed in three types of bacteriophage.

The most well characterized extrachromosomal prophage forming phage is P1. P1 is maintained as a circular, double stranded DNA, low copy number plasmid in infected cells that is spontaneously lost at very low frequency (10^{-5}) due in part to a phage encoded partition system and an addiction system (Lehnherr and Meyer 2000). The addiction system ensures that daughter cells contain a P1 plasmid by selectively killing cells that do not, based on the greater stability of a phage encoded toxin, Doc (Death On Curing), then the antidote Phd (Prevents Host Death) which are constitutively produced during lysogeny.

Bacteriophage N15 is unique among well characterized bacteriophages in that it is maintained extrachromosomally as a linear plasmid with covalently closed ends, the first such arrangement described in prokaryotes (Ravin et al. 2000, Svarchevsky and Rybchin 1984). N15 has a lambda like morphology and is maintained as a low copy number plasmid in infected cells (Ravin and Shulga 1970). The plasmid N15 replicated by prophage-encoded enzymes and is passed from mother to daughter cells nder the influence of a plasmid partitioning system most closely related to homologous systems in E.coli F factor and P1 and more distantly to bacterial chromosome partitioning systems (Grigoriev and Lobocka 2001).

Ff phages, Ff, fd, f1, M13, and CTX, are filamentous in morphology consisting of a tube of structural protein surrounding the single stranded phage genome. These coliphage infect by binding to the F-pilus of conjugating *Escherichia coli* causing the pilus to contract carrying the phage to the bacterial surface (Endemann et al. 1992,

Endemann et al. 1993). The entire phage particle is then translocated into the host cell through the pore created by the pilus. Once in the cell, the single stranded DNA genome is replicated by host cell machinery forming the double stranded, replicative form (RF) of the phage (Peeters et al. 1987). The RF phage is maintained extrachromosomally in low copy number and infected bacteria continually giving off progeny phage while they grow and divide normally. Two filamentous phage, CTX (Huber and Waldor 2002) and cf (Kuo et al. 1987), have been shown to also form lysogens by integration.

P4 is a coliphage that can be maintained as either an extrachromosomal plasmid (Deho et al. 1984, Lindqvist and Six 1971) or as a stably integrated prophage utilizing both aspects of the temperate lifestyle (Deho et al. 1984). This double stranded DNA phage exploits a helper phage, P2, to carry out lysis. Infection of a host cell that is not already carrying a P2 prophage always results in lysogeny since P4 does not carry any structural genes. 99% of the phage infecting non-P2 lysogens will integrate forming a P4 prophage in the host chromosome. However, P4 forms extrachromosomal prophages that are maintained as a low copy number plasmids in 1% of infections (Deho et al. 1984). Upon infection of a bacterial cell containing a P2 prophage, P4 can undergo lysogeny by integrating into the host genome, or it can undergo lytic growth using the structural genes of P2 create phage particles (Six 1963).

I.C. Site Specific Recombination

I.C.i. Conservative Recombination versus Transposition

Polynucleotidyltransferase proteins known as recombinases catalyze recombination by two methods: conservative recombination and transposition (Craig 2002). Conservative recombination usually occurs site-specifically and is carried out by

tyrosine and serine recombinases. No high energy or metal ion cofactors are utilized during conservative recombination, and no DNA is degraded or synthesized, so the total number of phosphodiester bonds remains unchanged (Figure 3 A and B). Transpostion involves the endonucleolytic cleavage of the target DNA, the insertion of a blunt ended transposon leaving gaps on either side of the insertion that are repaired by host DNA repair proteins (Figure 3 C).

During conservative recombination, four strands of DNA are cleaved and religated to each other by four recombinase subunits. The DNA substrates, or attachment sites, consist of symmetrically positioned protein binding sites flanking a short asymmetric crossover region where strand exchange occurs. The total size of the attachments sites depends on protein cofactors and secondary binding sites within the sequences and range in size from as few as 30 more then 400 base pairs. Serine recombinases create double stranded breaks within the target sequence by cleaving all four DNA strands before strand exchange occurs (Figure 3 B). Tyrosine recombinases cleave and religate pairs of strands at a time creating a Holliday Junction intermediate that is resolved by a second round of cleavage and ligation (Figure 3 A).

The mechanism of transpositional site-specific recombination is a two-step process (Figure 3 C). In the first step, an endolytic cleavage occurs between the 3′-OH of the transposon and the 5′-phosphate of the host on either side of the transposable element. The 3′-OH of the transposon carries out a nucleophilic attack on a phosphodiester bond at a new insertion site during the second step of the recombination reaction. The ends of the transposon are joined by transesterification to one of the two strands of the insertion sequence. These ends are separated by a 5′ staggered gap of 2-9 base pairs that will be filled by synthesis of the opposite strand or

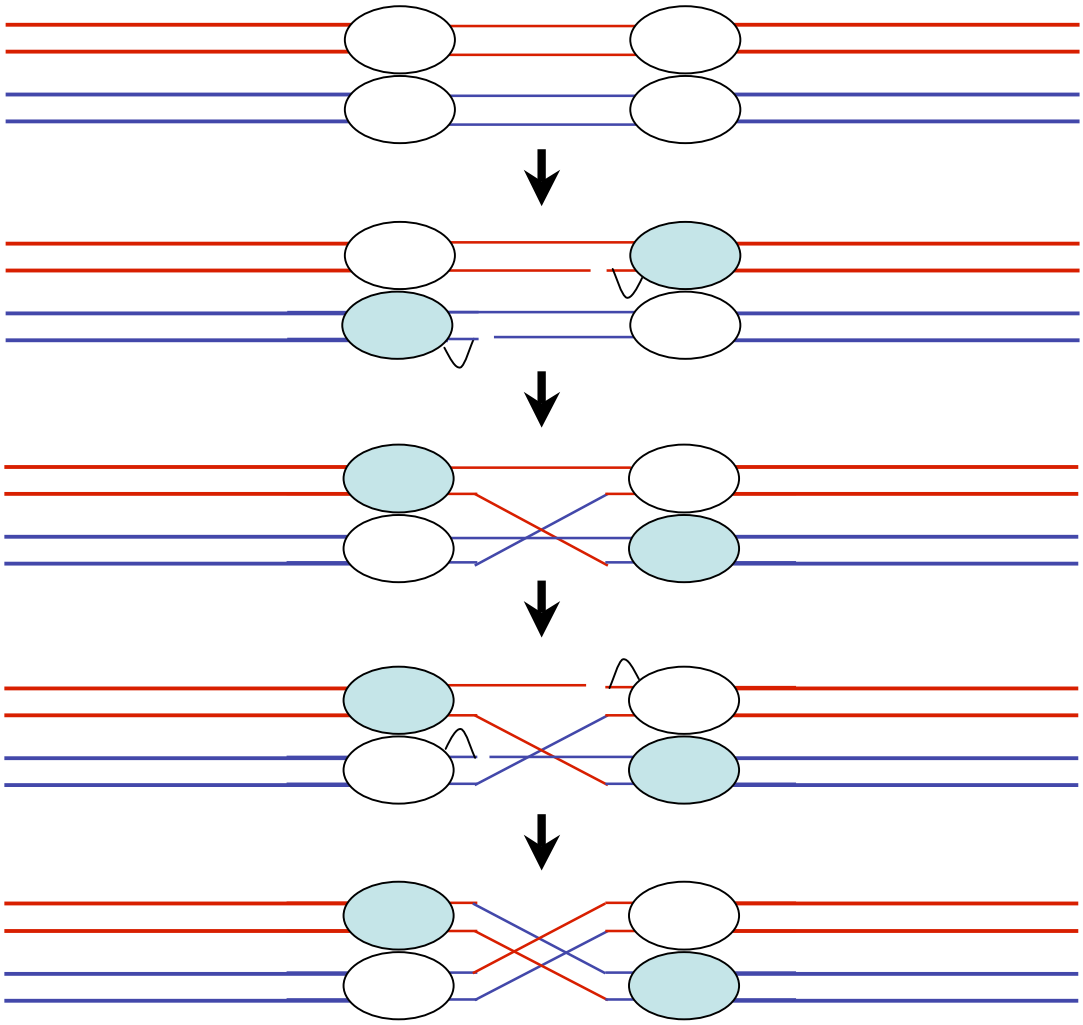
Figure 3. Conservative Recombination Versus Transposition

A. *Tyrosine Recombinase Mediated Conservative Recombination*. Four protomers of the tyrosine recombinase bind both DNA substrates at sites flanking a 6-8 base pair overlap region. Two protomers opposite each other are activated (light blue ovals) nicking the top strands of both DNA substrates and ligating each strand to its partner on the opposite substrate forming a Holliday Junction intermediate. An isomerization occurs which puts the Holliday Junction into position for the second round of strand exchange and activates the previously inactive protomers. These protomers nick and ligate the bottom strands completing the reaction.

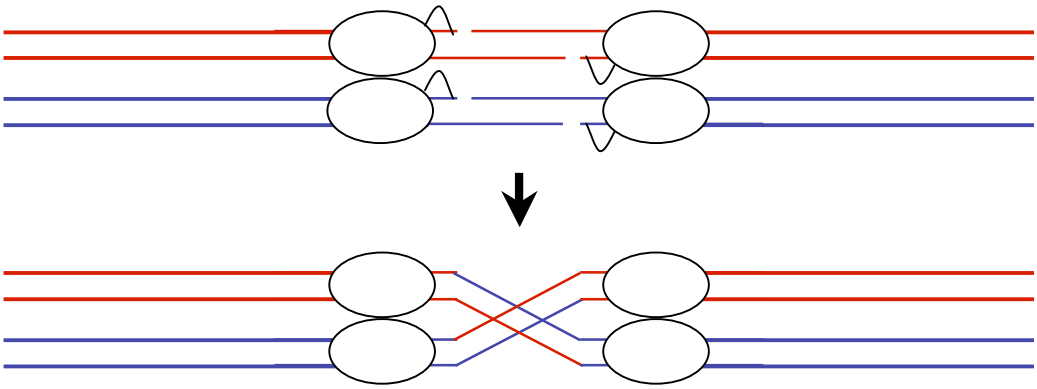
B. *Serine Recombinase Mediated Conservative Recombination*. Serine recombinases bind to a specific sites on the DNA substrates, and create double stranded breaks in both of the DNA substrates within a short, 2 base pair region. The strands are then exchanged and ligated to their partner DNA.

C. *Transposition*. Unlike the site-specific recombination described above, transposition usually occurs between segments of DNA at non-homologous sites. Protomers of the transposase bind to the target sequence and nick the DNA substrate on either side of a 2-9 base pair segment of DNA. Another set of protomers bind the free ends of the transposon DNA. The transposable element is then ligated to the free 3'OH of the top and bottom strand of the target leaving 2-9 base pair gaps on either side of the insertion. Gap repair enzymes (grey ovals) fill in the gaps leaving a repeat on either side of the transposable element.

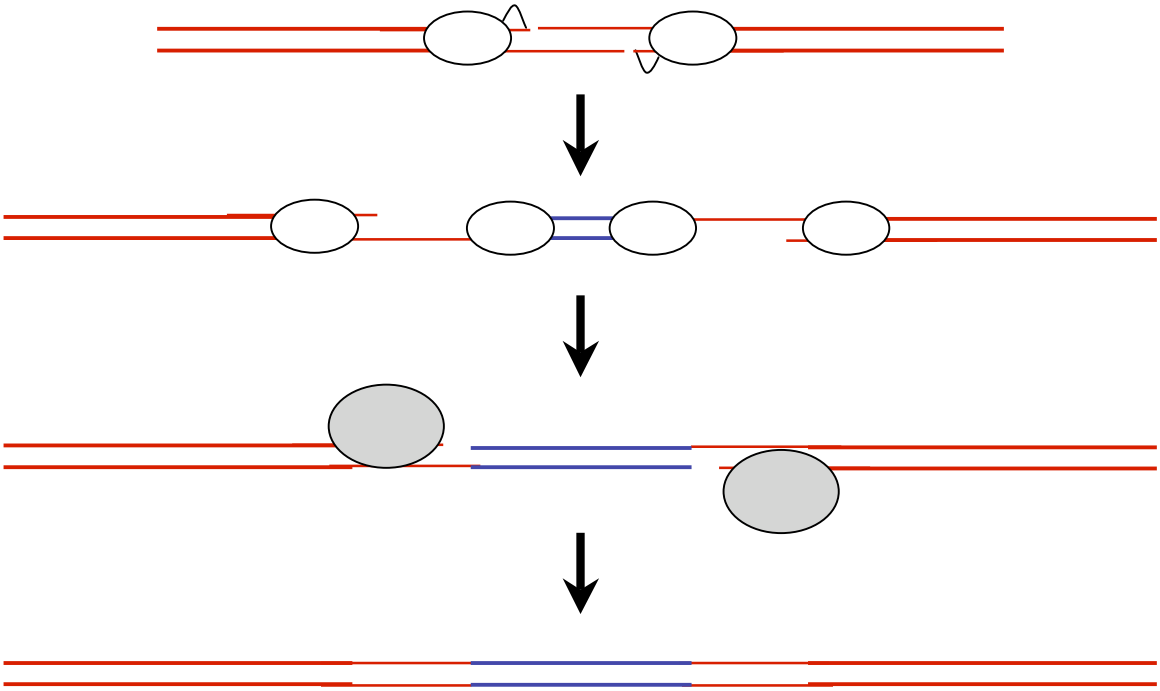
A



B



C



gap repair by cellular enzymes. Insertion of the transposon leaves a short duplication of the target DNA.

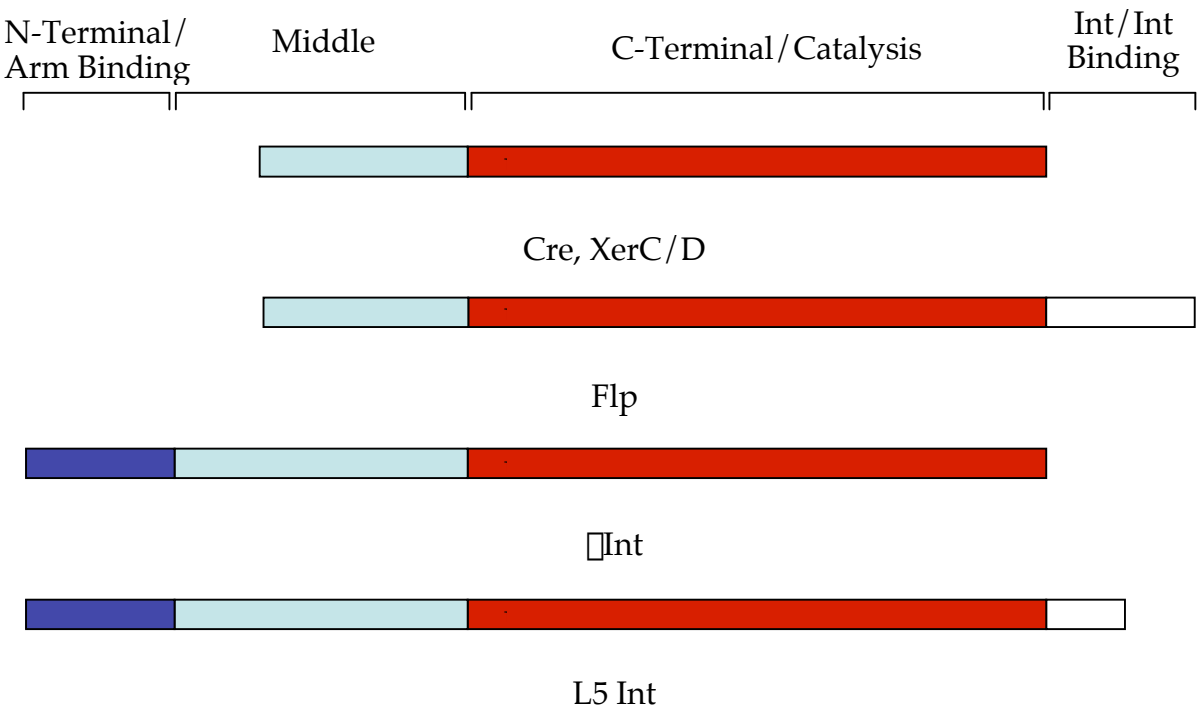
I.C.ii. Tyrosine Recombinases

Over 100 members of the tyrosine family of recombinases have been identified by sequence homology (Nunes-Duby et al. 1998). Members of this family of recombinases are found in diverse organisms from archea, eubacteria and bacteriophages to yeast and mitochondria and are closely related to type 1B topoisomerases both structurally and in the chemistry of their reactions (Cheng et al. 1998). The recombination reaction is carried out by four subunits of recombinase that simultaneously bind specific sequences within the substrate DNAs. The active site of these proteins is characterized by a strictly conserved arginine-histidine-arginine triad of residues and an active site tyrosine that carries out the nucleophilic attack on the DNA substrates (Argos et al. 1986). The active site tyrosine may be either part of a single active site, cis, as in most phage recombinases, or donated by an adjacent active site, trans, as in Flp recombinases (Chen et al. 2000, Lee et al. 1999). Strand exchange occurs one pair at time, so the reaction must be coordinated so that two subunits are activated while the other two subunits are inactive during the first round of cleavage and ligation. During the second round of cleavage and ligation, the opposite pair of subunits are activated while the previously active pair of subunits are inactive.

Tryosine recombinases can be divided into distinct domains (Figure 4) (Argos et al. 1986). The catalytic domain binds to the DNA substrates and catalyzes the recombination reaction (Hoess et al. 1990, Pan et al. 1991). A smaller domain is opposite the catalytic domain of some tyrosine recombinases that binds to arm type binding sites found in their DNA substrates (Tirumalai et al. 1997, Tirumalai et al. 1998). Cre, XerC/D, and Flp do not have arm type binding domains, and bind substrate through

Figure 4. Comparison of Tyrosine Recombinases

Tyrosine recombinases can be broken into 3 distinct domains. The catalytic domain (red) is at the C-terminal end of the protein and binds core where it catalyzes recombination. The catalytic tyrosine (Y) that carries out the nucleophilic attack on the DNA substrates, as well as two conserved arginines (R_I and R_{II}) and a histidine (H) are found in this domain. Flp and L5 Int have a small domain attached to the C-terminus of the catalytic domain involved in Int/Int interactions. The N-terminal domain (dark blue) binds to arm-type binding sites on the one of the substrates. Cre and Flp do not bind either substrate outside of the core, and do not have an N-terminal arm binding domain. The middle domain links the catalytic and arm binding domains (light blue). This domain has been implicated in core recognition.


N-Terminal/
Arm Binding
Middle
C-Terminal/Catalysis
Int/Int
Binding
Cre, XerC/D
Flp
Int
L5 Int

only the catalytic domain. A middle domain connects the two outer domains. This domain has been implicated in core recognition, but may perform other functions.

I.C.iii. DNA Substrates For Tyrosine Recombinases

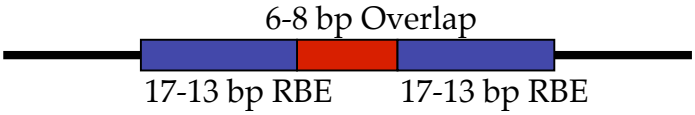
Tyrosine recombinases perform recombination within a 6-8 base pair asymmetric overlap region that must be identical in both substrates and is flanked on either side by recombinase binding elements (RBEs). RBEs are short, 7 – 13 base pair, sequences that usually contain inverted repeats, and recombinase subunits bind these sites while performing recombination on the overlap region. Although a few recombinases, Cre, Flp, and XerC/D, can perform recombination on substrates containing only RBEs and the overlap region, referred to as the core site, many substrates contain accessory sequence flanking the core (Figure 5).

Lambda recombination occurs at a 240 base pair recombination site (*attP*) that contains arm-type binding sites flanking the core. Protomers of λ Int bind arm and core sites of *attP* simultaneously, forming intramolecular bridges between these sites (Landy et al. 1979, Landy et al. 1980). A host-encoded DNA binding protein (IHF) binds *attP* specifically between arm and core binding sites, and catalyzes the bending of *attP* required to allow bivalent binding (Nash 1990). The phage-encoded excision protein (Xis) and another host-encoded DNA binding protein, Fis, also have binding sites between the core and arm-type binding sites.

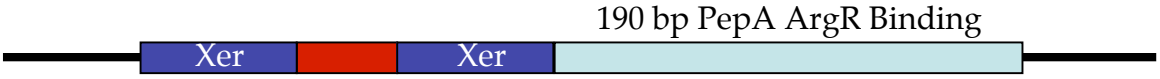
XerC/D recombination can occur at a 200 base pair recocombination site (*cer*) that contains the core flanked on one side by accessory sequence (Colloms et al. 1997). 190 base pairs to the right of core contain several binding sites for 2 accessory proteins ArgR, an arginine repressor, and PepA, a leucine amino peptidase, that are required for recombination (Alen et al. 1997, Colloms et al. 1997).

Figure 5. DNA Substrate for Tyrosine Recombinases

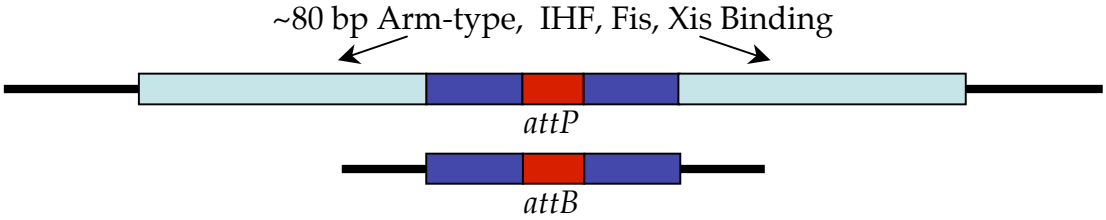
LoxP is the substrate for Cre recombinase and represents a simple recombination site that contains only a core site. The core contains a 6-8 base pair overlap region (red) flanked on either side by 17-13 base pair recombinase binding elements (RBEs)(dark blue). The substrate for XerCD recombination is Cer. Cer has an overlap region flanked by RBEs that bind XerC to the right, and XerD to the left. To the right of the XerC binding site is 190 base pairs accessory sequence (light blue) that bind PepA and ArgR. Binding of PepA and ArgR help produce the correct DNA topology for XerD binding. Cer sites are usually found in tandem. λInt performs recombination on different *attP* and *attB* substrates. Eighty base pairs of accessory sequence flanks a normal *attP*. The accessory sequence of λ*attP* is bound by Int at arm-type binding sites and contains binding sites for Integration Host Factor (IHF), another host factor FIS, and a phage-encoded excisionase (Xis). IHF, Fis, and Xis bind to and bend the accessory DNA so λInt can bind *attP* at arm-type and core binding sites.



LoxP Type Recombination Site



Dif Type Recombination Site



Lambda *attP*/*attB* Type Recombination Site

The bacterial attachment sites (*attB*s) for these recombinases contain only the core, so they are much smaller then the *attP*s. *attB*s do not contain accessory DNA, and *attB*s appear to be bound in the absence of host factors. *attB* capture is poorly understood, but probably occurs late in the formation of a recombinationally active complex.

I.C.iv. Structural Elements of the Catalytic Domain of Tyrosine Recombinases

The structures of 5 tyrosine recombinases have been solved; the catalytic domain of lambda (Kwon et al. 1997) and HP1 (Hickman et al. 1997) integrases, XerD recombinase (Subramanya et al. 1997), and the complete structures of Cre and Flp . Cre and Flp have also been crystallized in complex with their DNA substrates (Gopaul et al. 1998, Guo et al. 1997, Guo et al. 1999, Rice 2000). Structural information can be used to begin to understand how the recombinagenic complex is built on the substrate DNA, and how site-specific recombination occurs within the 6-8 base pair overlap.

There are extensive protein/DNA interactions at the interface between the recombinase subunits and the RBEs that typify the tight binding of the recombinase to core (Gopaul et al. 1998, Guo et al. 1997, Guo et al. 1999, Rice 2000). However, the four-subunit complex is held together by protein/protein interactions that occur in the same domain of the recombinase that binds core. As the recombinase's subunits bind to core, they also bind to neighboring subunits, creating intricate networks of interactions between the recombinase's subunits and the DNA substrates that hold the recombinagenic complex together and tether it to the substrates.

Intersubunit contacts also occur in the extreme C-terminus of the recombinase subunits. In many tyrosine recombinases, the most C-terminal helix of the one subunit is buried in the acceptor pocket of the adjacent subunit, and this subunit in turn buries its C-terminal helix in the subunit adjacent to it binding four recombinases subunits

Figure 6. Crystal Structures of Cre Recombinase

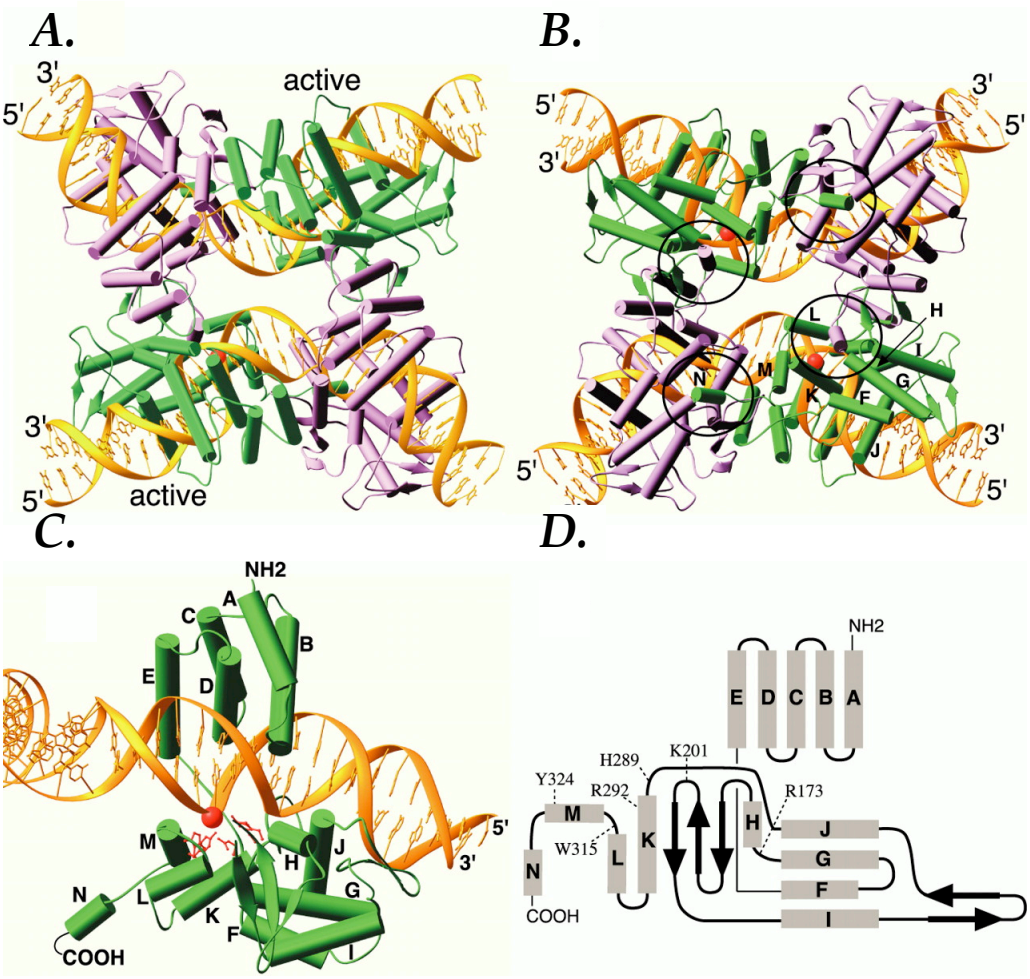
A. Four protomers of Cre bind to *LoxP* in Cre/*LoxP* co-crystal structures. This structure is arranged so that the N-terminal domain of Cre is facing up. Active protomers are indicated and shown in green, and Inactive protomers are shown in purple. The DNA is shown in yellow and is bent severely in the active site of Cre. Close examination also shows that protomers on the same substrate DNA have a more flexible interface, whereas protomers on opposing substrates have a tighter interface.

B. Four protomers of Cre are bound to *LoxP* and are viewed from the C-terminus in this panel. Active protomers are green, and inactive are purple. The DNA is yellow. The helices of the lower left active proteomer are labeled F through N. Circles are provided to highlight interactions between helix N and the neighboring subunit.

C. A single active proteomer of Cre is shown in green binding to *LoxP* shown in yellow. The carboxyl and amino termini are shown, and the helices are labeled A-N. The active site tyrosine as well as arginine/histidine/arginine triad of coordinating residues are shown in red with the sissile phosphate, the red ball on the substrate DNA. Cre makes C-shaped clamp around *LoxP* with the N-terminal domain binding in the major groove opposite the sisile phosphate. The M helix, which contains the tyrosine nucleophile, is in position for nucleophilic attack on the sissile phosphate.

D. A linear representation of a single Cre proteomer is shown here. □-helices are shown as grey bars, and □-sheets are shown as black arrows. Black lines

represent random coils and turns. Conserved residues of the catalytic domain are shown with their amino acid numbers.



(Van Duyne 2001)

together cyclically (Figure 6 B) (Abremski and Hoess 1984). Flp depends on the exchange of helices for catalysis. The helices exchanged during the building of the Flp recombination complex contain the catalytic tyrosine indicating that recombination is occurring in trans. Therefore, the active subunit's nucleophilic attack is carried out utilizing the catalytic tyrosine of the adjacent subunit (Chen et al. 2000, Lee et al. 1999). Not all recombinases perform a cyclic exchange of C-terminal helices. HP1 subunits exchange the helices reciprocally demonstrating that interactions between subunits at the extreme C-terminus can occur in a variety of ways (Hickman et al. 1997). The exchange of helices between subunits creates a complex that is held together tightly at the active site creating a nearly 4 fold symmetric complex. This complex is not affected by substrate binding since substrate binding occurs elsewhere in the catalytic domain.

The catalytic domain binds to the substrate DNA by forming a C-shaped clamp around the substrate (Figure 6 C) (Guo et al. 1997). The N-terminus of the catalytic domain binds to the major groove of the DNA substrate proximal to the crossover region. The C-terminus of the catalytic domain contains the catalytic triad, R-H-R, two histidines (HII and H/WIII, HIII is a tryptophan in several recombinases) and a lysine (K□) that have been shown by mutagenesis to be necessary for recombination (Cheng et al. 1998) and interacts with the DNA on the opposite face of the substrate. The R-H-R triad coordinates the scissile phosphate on the substrate DNA (Grainge and Jayaram 1999). HII is a proton acceptor during the cleavage part of the recombination reaction and donates the proton during the ligation step (Guo et al. 1997, Pan et al. 1991, Sherrat and Wigley 1998). H/WIII binds to the scissile phosphate via their indole-N. K□ is in a largely unstructured loop connecting □-strands 2 and 3 (Cao and Hayes 1999). K□ helps to organize the active site by binding to the base pair adjacent to the cleaved phosphate

in the minor groove of the substrate DNA it then activates the 5′-OH of the leaving group (Guo et al. 1999, Redinbo et al. 1998). The active site tyrosine is located in the penultimate helix (M) (Figure 6 D). Helix M is flexible in unbound structures and moves in and out of the active site parallel to the helical axis (Van Duyne 2002). When substrate is bound and the recombinagenic is properly formed, helix M moves into position for the nucleophilic attack indicating the active site is assembled on the substrate.

When bound to DNA, the four subunit recombinagenic complex has strict dyad symmetry implying that the DNA substrates are arranged in an antiparallel conformation. This arrangement could be necessary for strand exchange without large structural perturbations in the recombinase subunits since the DNA substrates in co-crystal structures are severely bent (Figure 6 A). *LoxP*, the Cre recombination site, has a 75^{o} bend in the overlap region when in complex with Cre (Guo et al. 1999), and Flp introduces a 140^{o} bend in frt, it's target DNA. For a recombinagenic complex to introduce such a severe bend in the active site DNA, interactions between the substrate core and recombinase subunits must be strong. Intersubunit interactions must also be solid, so the complex does not come apart during recombination (Luetke and Sadowski 1998, Schwartz and Sadowski 1990).

I.C.v. Chemistry of the Recombination Reaction

Once the recombinagenic complex has been built on the substrate DNA and the active site has been assembled, the chemistry of recombination can proceed (Sherrat and Wigley 1998). The two recombinase subunits on opposite corners of the of the four subunit recombinagenic complex are activated in the first chemistry step. These subunits cleave the top strand of the substrate DNA at 5′ end of the crossover region. The tyrosine nucleophile forms a 3′-phosphotyrosine linkage leaving a free 5′-OH that

then attacks the 3′-phosphotyrosine of the partner strand forming a Holliday Junction (Figure 7).

The second step in the chemistry of recombination requires the movement of the substrate into position for the second round of cleavage and ligation. Two processes, branch migration and strand swapping, have been implicated in the rearrangement of the DNA substrate.

During branch migration, the branch point in the substrates resulting from the first round of cleavage and ligation moves across the entire length of the crossover region to a point close to the base pair involved in the second nucleophilic attack (Figure 8 A)(Weisberg et al. 1983). This would require significant movement of the protein subunits since the structure of the DNA would change significantly during branch migration, but the recombinase subunits would be required to maintain specific contacts with the substrate DNA (Stewart et al. 1998).

More recent structural data suggests that branch migration may not be the method of substrate rearrangement (Nunes-Duby et al. 1995). X-ray crystal structures of Cre intermediates suggest that the branch point is close to the middle of the overlap region (Figure 8 B) (Gopaul and Duyne 1999, Gopaul et al. 1998, Guo et al. 1997). The structure shows helical restacking of the DNA substrates with the exchange of stacking partners and limited branch migration. The Holliday Junction that results from the first round of cleavage and ligation is in a square planar conformation and the basepairs are all unstacked (Duckett et al. 1988). Watson Crick base pairing occurs between the first 2 adjacent base pairs of the partner substrates. This might allow the recombinase to sense the homology of the partner sequences in the crossover region since mismatches would create stereochemical barriers for ligation that would stop the recombination reaction from going to completion. Restacking of the substrate DNA would also limit the

Figure 7. Recombination Chemistry

In panel one, the active site tyrosine attached to a globular recombinase (dark blue half oval) is attacking the phosphate (light blue circle) of a single strand of DNA. DNA bases are shown in red boxes facing away form the recombinase. The OH of the active site tyrosine attacks the 5′ phosphate of the lower nucleotide. Forming a 5′ phosphotyrosine bond shown in the second panel. The 3′ OH of the upper nucleotide attacks the phosphotyrosine bond between the partner DNA strand and the other active recombinase proteomer shown at the bottom of panel 3. The new phosphodiester bond has been formed in the final panel, and the active recombinase protomers are made inactive as the tyrosine is withdrawn from the active site.

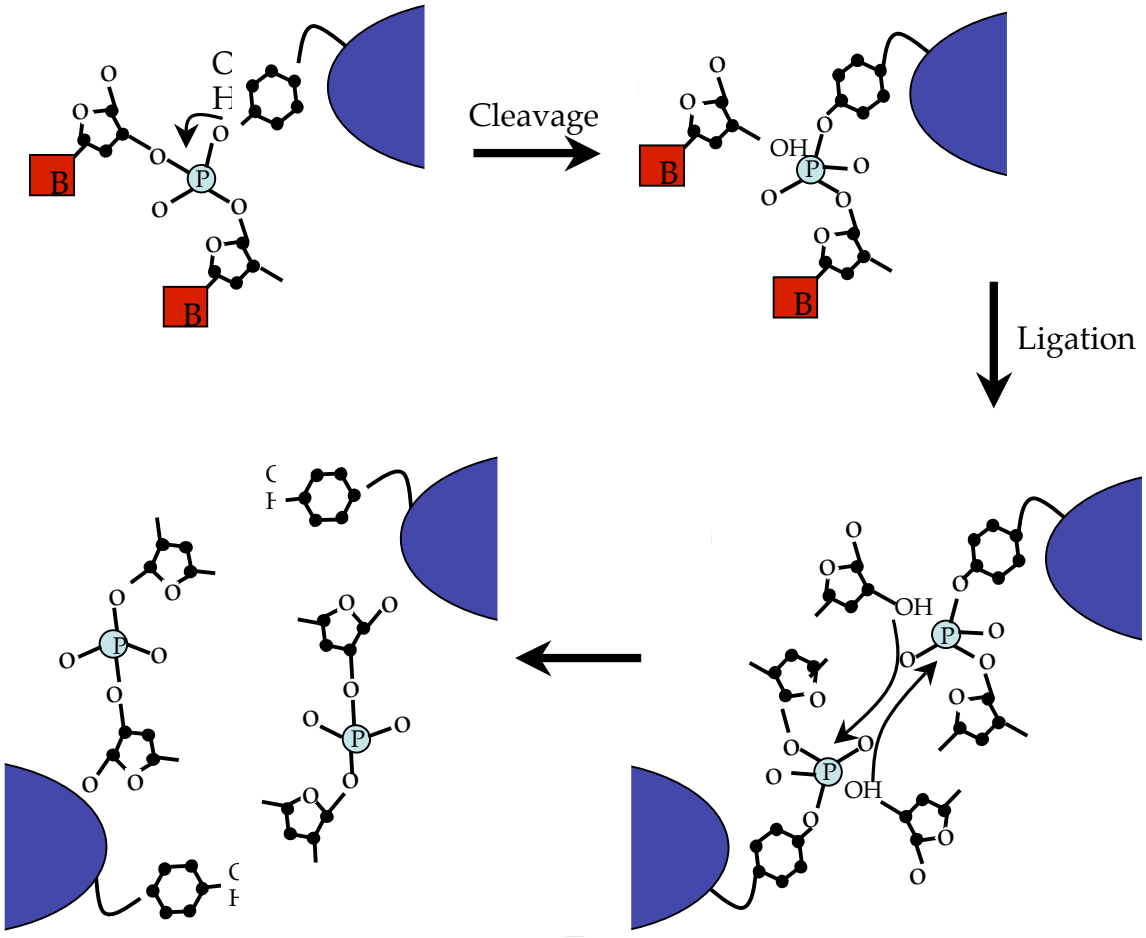
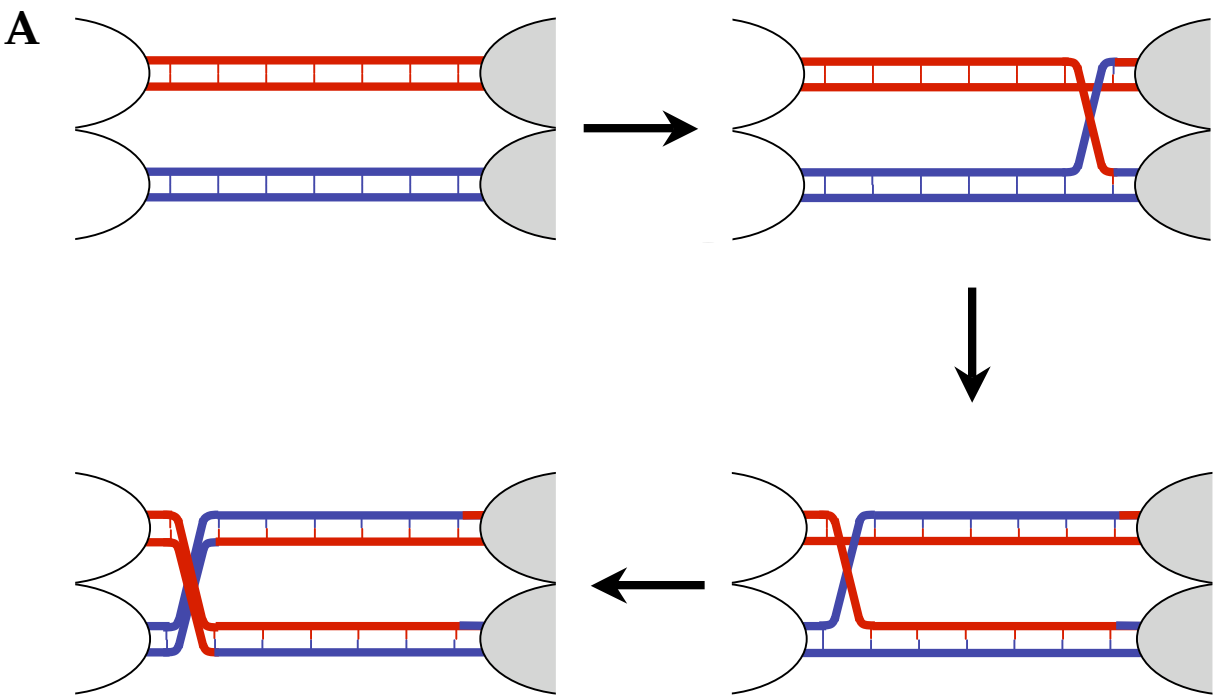
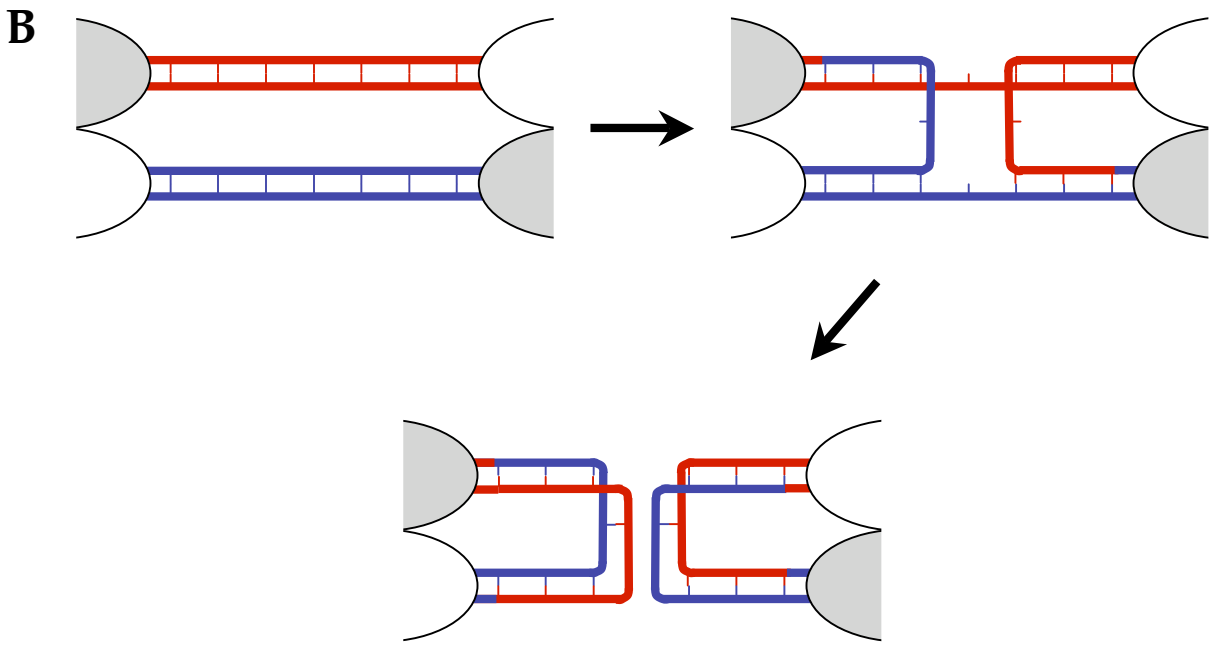

Cleavage
Ligation
B
P
O
OH

Figure 8. Branch Migration and Strand Swapping

A. Four protomers of recombinase (half ovals) are shown binding DNA substrates (read and blue) in the first panel. Two of the protomers are activated (grey half ovals) and catalyze the first round of strand exchange leaving a Holliday Junction (red and blue X). During isomerization, Holliday Junction is moved across the active site and closer to the inactive protomers. The exchanged strands pair with their partner strands and the homology of the overlap region is tested. An unpaired base pair stops recombination from going to completion, and the reaction is reversed. Once in position, the second round of strand exchange occurs, and the recombination reaction is completed.

B. As above, four protomers of recombinase are bound to the substrate DNA in panel 1. The first round of strand exchange is carried out by opposing protomers (grey half ovals) leaving a Holliday junction that is in the middle of the active site several base pairs have been paired with the partner strands DNA allowing limited homology sensing and leaving several unpaired bases on both strands. The inactive protomers are then activated, and the second round of strand exchange is carried out resolving the Holliday Junction.


A


B

amount of movement required in the recombinase subunits following isomerization, so the protein could achieve the proper conformation for the second strand exchange without large structural purturbations. Isomerization is coupled to the rearrangement of the DNA substrates and involves a 17^{o} rotation of the recombinase subunit that includes minor remodeling of the inter subunit contacts (Nunes-Duby et al. 1995).

In the final step of the recombination process, the Holliday Junction is resolved. Holliday Junction resolution involves similar chemistry to the first nucleophilic attack on the substrate DNA, and is carried out by the pair of recombinase subunits that were inactive during the first round of cleavage and ligation. The active site tyrosine perfoms a nucleophilic attack on the bottom strands of the Holliday Junction allowing the strands to be fully exchanged and the product, a single DNA molecule containing both substrates is released. Some data suggests that the subunits remain attached to the product for some time following the reaction (Peña et al. 1999).

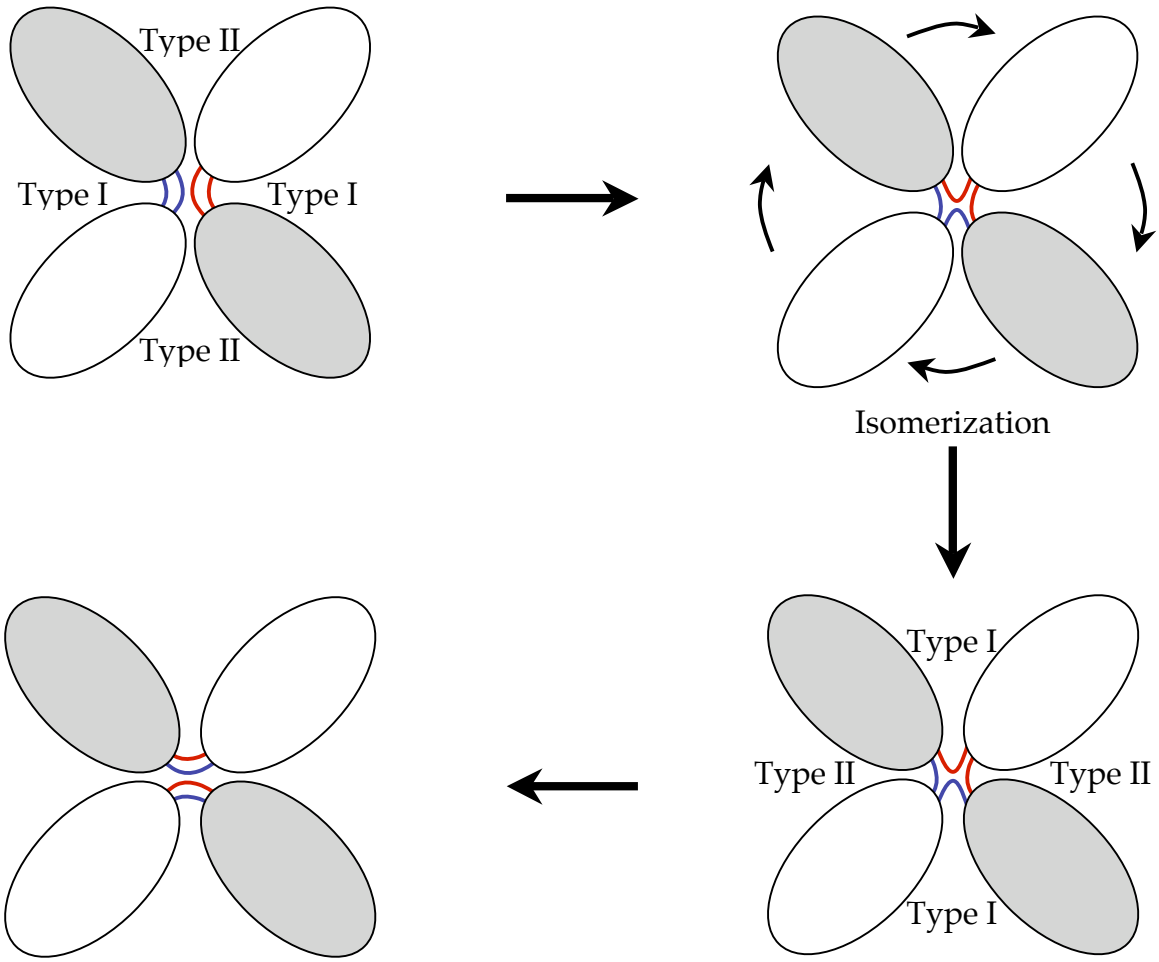
I.C.vi. Regulation of the Active Site

For tyrosine recombinases to carry out recombination using the chemistry as described above, the active site must be highly regulated. Only 2 of the 4 subunits can be active at a time for the reaction to occur, and halfway through the reaction, the active subunits must switch (Arciszewska et al. 2000, Hallet et al. 1999). This would require 2 types of interface. The type I interface is created between the 2 subunits bound to the same DNA molecule. The type II interface is formed between the 2 subunits bound to different DNA strands (Gopaul and Duyne 1999). Therefore, each subunit is bordered by one of each type of interface, and the interfaces type switches when isomerization at the Holliday Junction intermediate (Figure 9).

There are differences between type I and type II interfaces in at least two crucial areas. Structural data shows that helices M, which carries the catalytic tyrosine, and N,

Figure 9. Isomerization and Int/Int Interfaces

Four protomers of recombinase (ovals) are bound to the DNA substrates (red and blue curved lines). Active protomers (grey ovals) are bordered by type I interfaces with protomers bound to the same DNA substrate and type II interfaces with neighboring protomers bound to the other substrate. The protomers of the type II interface are more tightly associated, and are shown closer together then the protomers of the type I interface. After the first round of strand exchange, an isomerization activates the previously inactive subunits and changes the interfaces between protomers. The previously type I interfaces are now type II and the type II interfaces are now type I. After the isomerization the second round of strand exchange occurs, and the Holliday junction is resolved.


Type II
Type I
Type I
Type II
Isomerization
Type I
Type II
Type II
Type I

which is donated to the adjacent subunit (Figure 4 D), are more extended in the type I interface. In the type II interface, the region between the M and N helices is much less flexible. The N helix is donated from an inactive to an active subunit, and the linker region between helixes M and N is more compact even though it is not part of the active site. It is thought that flexibility in this region of the type I interface is necessary to allow helix M to be positioned for nucleophilic attack of the tyrosine on the substrate DNA (Gopaul and Duyne 1999, Gopaul et al. 1998).

The opposite is true for β-strands 2 and 3 (Figure 4 D). These strands are more flexible in the type I interface than the type II interface. β-strands 2 and 3 are positioned to give lysine-β access to the active site in the active, type II interface. β-strand 3 contacts helices M and N's linker indicating that the β-strands 2 and 3 region may be involved in aligning active site tyrosine (Gopaul and Duyne 1999, Gopaul et al. 1998).

By alternating the flexibility of a few key structural elements, the tyrosine recombinase can control the activity of recombinase subunits that make up the recombinagenic complex. Presumibly, rearrangements in the DNA structure caused by the catalysis of recombination trigger changes in the recombinase structure, so structural rearrangements in the DNA and recombinase subunits can be tightly coupled.

I.C.vii. Cooperativity in Tyrosine Recombinase Complexes

Most of the structural data that exist are based on the crystal structures of a few simple recombinase systems that do not require bivalent binding to substrate DNA or interactions with host factors. However, many tyrosine recombinases require more elaborate protein/DNA structures that require binding to arm-type sites on substrate DNA and interactions with other DNA binding proteins. The integration of phage

lambda is a good example of a recombination system that requires more elaborate protein/DNA interactions.

The phage attachment site of lambda Int (λ Int) contains the core flanked by arm-type binding sites P1 and P2 to the right and P'1, P'2 and P'3 to left (Better et al. 1982). IHF, a host encoded DNA binding protein binds site-specifically to the region of DNA between P2 and core called H2 (Nash 1977, Nash 1981). Mutagenesis of H2 eliminates the ability of IHF to bind to *attP* DNA in the absence of λ Int. However in the presence of λInt, IHF can bind to *attP* in this region of *attP* and create the specific bend in *attP* necessary for recombination to occur. This has been attributed to pseudo-specific binding of IHF that could be mediated by protein/protein interactions between IHF and λInt (Segall et al. 1994) indicating that there are cooperative interactions between λInt and IHF that regulate the positioning of IHF between P2 and core.

IHF introduces a sharp bend in *attP* DNA (Rice et al. 1996, Thompson and Landy 1988) that is thought to be important in bringing the core and arm-type binding sites into close enough proximity for bivalent binding (Better et al. 1982, Kim and Landy 1992, Richet et al. 1988). In the same way as IHF recognizes a specific site between P2 and core where it binds *attP*, λXis recognizes two 13bp imperfect repeats between P2 and core of *attR* (X1 and X2) (Abremski and Gottesman 1982, Yin et al. 1985). Xis is thought to catalyze *attR* bending allowing the formation of intramolecular bridges between arm-type and core binding sites that are necessary for the formation of a recombinationally active excision complex (Better et al. 1983). Some evidence suggests that protein/protein interactions between Xis and Int may play an important role in activating recombination (Thompson et al. 1987a) indicating again the importance of cooperative interactions in the recombinagenic complex of λInt.

The carboxyl terminal domain of λXis is competent to stimulate formation of a λInt/*attR* complex, but does not appear to bind X1 or X2 of *attR*. In contrast, the amino terminal domain of λXis binds *attR* DNA but is not sufficient to stimulate excisive recombination (Numrych et al. 1992, Wu et al. 1998). In experiments using DNA substrates consisting of a single arm-type binding site (P2) with X1 and X2, λInt alone does not form an electrophoretically stable complex. However, the presence of λXis stimulates the formation of an Int-Xis-DNA complex that is formed cooperatively. λInt truncation mutants show that the amino terminal 64 amino acids of λInt are competent to make these interactions (Cho et al. 2002). These results indicate that cooperative protein/protein interactions between Xis and Int might modulate the binding activity of the amino terminal arm-type binding domain of Int and, therefore, the activity of Int itself.

Cooperative interactions have also been observed between the arm-type binding domains of λInt alone. The N-terminal 65 amino acids of λint are competent to bind arm-type binding sites of λ*attP* non-cooperatively. However, cooperative interactions can be established with arm-type binding sites when truncated λInts contain the first 70 amino acids (Sarkar, Azaro et al. In press) indicating that interactions in the amino terminal domain allow int protomers to bind the arm-type sites of *attP* cooperatively.

The importance of Int binding to arm-type binding sites are evident in the activity of the C-terminal catalytic domain when separated from the arm-type binding domain. The C-terminal domain of λInt retains topoisomerase activity in the absence of the amino terminal domain and actually binds to *attP* core containing DNA with higher affinity. Addition of the N-terminal domain in trans stimulates the ability of the C-terminal domain to cleave core and, at the same time, lowers the affinity of the C-

terminal domain for core (Sarkar et al. 2001). Conversely, Ints with mutations in the core binding domain are also defective in binding to arm-type sites (Han et al. 1994a). These results suggest that the bivalent binding of int is important in regulating its activity and that the lines of communication between these domains go in both directions.

I.D. Mycobacteriophage L5

I.D.i. General Background

Mycobacteriophage L5 was isolated from a lysogenic strain of *Mycobacterium smegmatis* in 1960 (Doke 1960). L5 morphologically consists of an icosahedral head and a long flexible tail. Tail fibers are abscent, but there appears to be foot-like swelling at the base of the tail (Hatfull and Sarkis 1993). The genome of L5 consists of 52,297 base pairs (Hatfull and Sarkis 1993), and the G+C content of L5 is similar to its host *Mycobacterium smegmatis* 63.2% (Clark-Curtiss et al. 1990). There are a total of 85 putative protein-coding regions and 3 tRNA genes in the linear genome.

The phage attachment site (*attP*) divides the L5 genome of L5 in half by. Unlike lambda whose genome contains divergent promoters (Ptashne et al. 1976), the right arm of L5 is transcribed to the left toward *attP* by a single promoter found between genes 88 and 1, P_{Left} (Nesbit et al. 1995). A promoter for the left arm has not been identified. The left arm contains the structural and lytic genes that are repressed during lysogenic infection by the product of gene 71 (gp71). Gp71 is a small DNA binding protein that represses P_{left} and binds to 23 novel stoperator sites scattered throughout the L5 genome. The binding of gp71 to stoperator sites represses gene expression by limiting the elongation of gene transcript (Brown et al. 1997).

I.D.ii. Lysogeny of Mycobacteriophage L5

L5 lysogens contain a single integrated copy of the L5 genome. The bacterial attachment site, *attB*, is located within the 3′ end of $tRNA^{Gly}$ (Lee and Hatfull 1993). Integration occurs at the left end of a 43 base pair region that is shared between *attP* and *attB* within a 7bp overlap region where strand exchange takes place (Peña et al. 1996). *attP* and *attB* as well as recombination product *attL*, must contain the 43 base pair common region since it is necessary for the reconstitution of the essential $tRNA^{Gly}$ gene (Lee et al. 1991).

Site-specific integration occurs through the action of a phage encoded integrase (L5 Int), the product of gene 33, which belongs to a large family of tyrosine recombinases. Together with a host encoded DNA binding protein mycobacterial Integration Host Factor (mIHF) (Pedulla et al. 1996), L5 Int binds to *attP* and inserts the phage genome into the host chromosome at *attB* (Lee and Hatfull 1993). This process is reversed in the presence of the phage encoded excisionase (Xis-L5), the product of gene *36*, which appears to facilitate the formation of a stable *attR* complex by binding to L5 *attP* between the arm-type binding site P2 and the core (Lewis and Hatfull 2001).

I.D.iii. Protein Requirements for L5 Int Mediated Recombination

L5 integrase (L5 Int) is a member of the tyrosine family of recombinases and contains the active site Tyrosine (Y342) along with the Ariginine, Histidine, Arginine (R-H-R) motif common to this family. Int can be divided into 2 major domains that are proteolytically sensitive at their domain boundary. The carboxyl terminal catalytic domain contains the catalytic residues and binds to *attP* core and *attB*. The amino terminal domain is much smaller then the carboxyl terminal domain consisting of the first 57 amino acids and binds to the arm type binding sites on *attP*. A small carboxyl terminal tail is thought to be necessary for protein/protein interactions between Int

subunits (Abremski and Hoess 1984, Lee et al. 1991). The region between the C- and N-terminal domains, the middle domain, connects the two outer domains, but little is known about this region (Figure 10 A).

During the recombination reaction, four units of Int are thought to bind to the core and arm type binding sites forming a higher order protein/DNA complex. Recombination occurs within the 7 base pair overlap sequence within the core-binding region using chemistry similar to other tyrosine recombinases with Y342 performing the nucleophilic attack on the substrate DNA. The reaction occurs in the absence of any co-factors including magnesium or ATP. However, spermidine is required for integration *in vitro* (Lee and Hatfull 1993).

A host encoded DNA binding protein, mIHF, mediates the bending of the substrate DNA between arm-type binding sites and the core (Pedulla et al. 1996). Int mediated recombination requires mIHF, and other DNA binding proteins that are known to stimulate recombination of λInt, IHF or HU, can not be substituted in the reaction (Lee and Hatfull 1993).

I.D.iv. DNA Substrates for L5 Integrase Mediated Recombination

The L5 *attP* site is composed of a core binding site consiting of a 7 base pair overlap region flanked on either side by 7 base pair RBEs containing imperfect inverted repeats in which every other base pair is part of the inverted repeat (Peña et al. 1997). Seven 10 base pair arm-type binding sites (P1-P7) are found in L5 *attP* on either side of the core (Figure 10 B). These sites were initially identified based on sequence similarity and subsequently shown to be bound by L5 integrase during recombination (Lee et al. 1991). To the left of core is a pair of arm type binding sites P1/P2 separated by a single base pair and a single site P3 which has been shown to be dispensable during both integration and excision (Lewis and Hatfull 2001, Peña et al. 1997). To the right of core

Figure 10. Protein and DNA Requirements for L5 Int Mediated Recombination

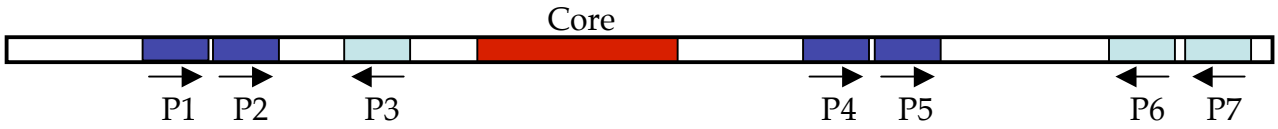
A. L5 Int is a tyrosine recombinase that can be divided into 4 distinct domains. The catalytic domain (red) contains the active site tyrosine (Y348) and the triad of catalytic residues (R205, H313, and R316). The extreme c-terminus (light blue) of the catalytic domain is thought to be necessary of Int/Int interactions. The N-terminal domain (dark blue) catalyzes binding to the arm-type sites of *attP*. Between these domains is the less well-defined middle domain (white).

B. L5 *attP* is composed of a core-binding domain where strand exchange occurs flanked on either side by 7 arm-type binding site (P1-P7, blue). Arm-type binding sites are arranged in pairs save P3, as shown, and share sequence similarity. P1/P2 and P4/P5 (dark blue) are utilized during integration and are all in the same orientation. P3 and P6/P7 remain unbound during integration and excision and are in the opposite orientation of P1/P2 and P4/P5.

A.


R205
H313
R316
Y348
N
Arm
Middle
Catalytic
C

B


Core
P1
P2
P3
P4
P5
P6
P7

are P4/P5 that are separated by one base and P6/P7 that are separated by two base pairs. L5 Int recombination can be carried out on substrates containing only P1/P2 and P4/P5 with equal efficiency to full length *attP* indicating that P6/P7 and P3 are not necessary for recombination (Peña et al. 1997).

The bacterial attachment site (*attB*) is much simpler than *attP*. A 7 base pair overlap region that is identical to the overlap region of *attP* makes up *attB*, and the overlap region is flanked on either side by RBEs with the same imperfect inverted as *attP*. The right side RBE is identical to the right side RBE of *attP*, however the left side RBE contains only three base pairs of the inverted repeats and shares 4 of the seven base pairs with the left side RBE of *attP*. There are no arm-type binding sites on *attB* and *attB* binds to L5 Int in the absence of accessory proteins (Peña et al. 1996).

The products of recombination are *attL* and *attR*. To the right of the integrated genome is *attR* that contains arm-type binding sites P1/P2 and P/3. The recombinagenic complex releases *attR* following integration. The left recombination product, *attL*, contains P4/P5 and P6/P7 and remains bound by L5 Int after recombination (Peña et al. 1997).

Between P3 and core to the left and between P4/P5 and core to the right, *attP* is bound by mIHF (Pedulla and Hatfull 1998, Peña et al. 1997), and Xis-L5 between P1/P2 and P3 during excision (Lewis and Hatfull 2001). There is very little sequence similarity between the regions where mIHF binds, so mIHF is thought to bind *attP* non-specfically (Peña and Hatfull 2000). Xis-L5 binds cooperatively to four sites (X1-4) that are found between P1/P2 and P3 in *attR* and share some sequence similarity (Lewis and Hatfull 2001).

I.D.v. L5 Integrase Complex Formation

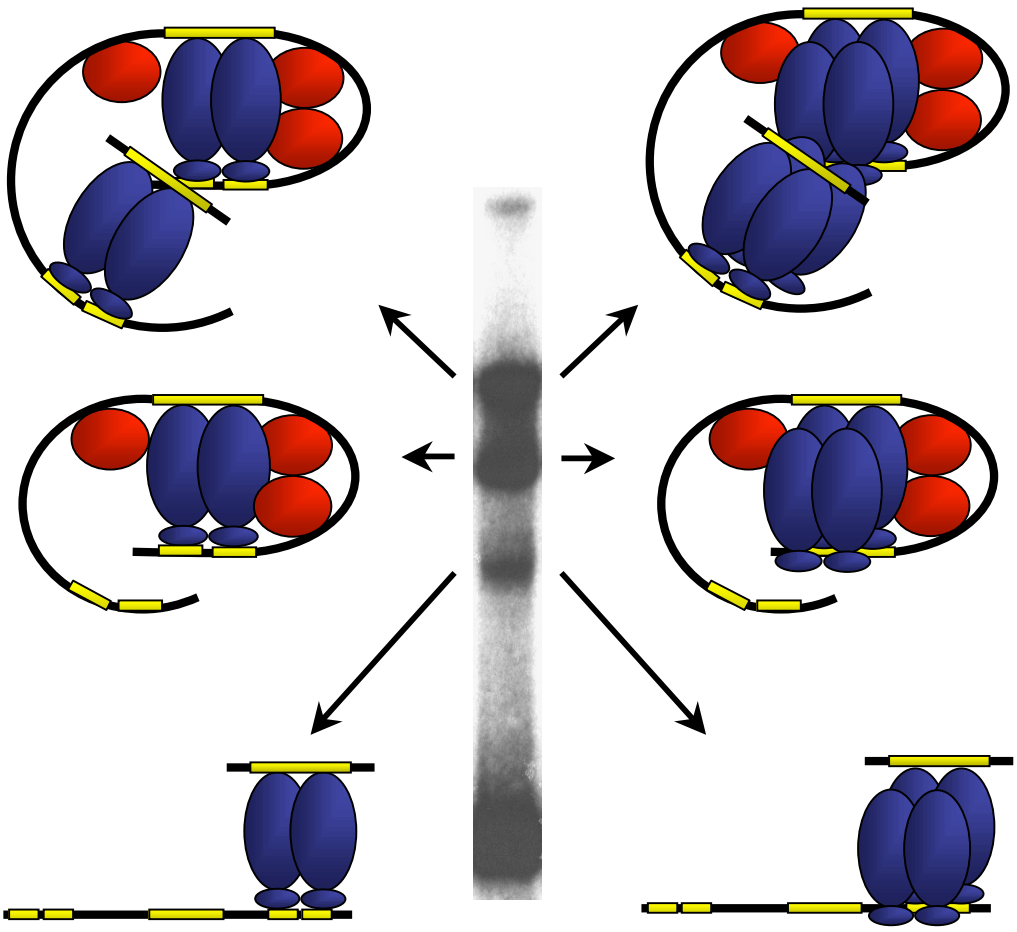
Several complexes have been identified and characterized using linear substrates on ice. These complexes can be separated using native gel electrophoresis, and DNAse I footprinting was used to determine which sites on *attP* are occupied by L5 Int binding (Peña et al. 1999) (Figure 11). The intasome was the first complex isolated, and this complex can be isolated at room temperature, as well as, on ice. Int appears to be forming an intramolecular bridge between core and P4/P5 in intasome, and this complex does not contain *attB* since it can be formed in the absence of *attB*. P1/P2 are unoccupied in this complex. Recombination can be accomplished by this complex if mIHF and *attB* are added to intasomes that are formed within the gel (Peña and Hatfull 2000).

Synaptic Complexes 1 and 2 are formed when *attB* is added to the reaction and can also be isolated using native gel electrophoresis. Synaptic Complex 1 contains Int forming intramolecular bridges between core and P4/P5 and intermolecular bridges between *attB* and P1/P2. Addition of mIHF to this complex allows recombination to be performed by this complex if mIHF is added to the complex. It is thought that this is an open complex, so mIHF is necessary to bend *attP* between P1/P2 and core placing the L5 Int protomers into position for recombination (Peña and Hatfull 2000). Synaptic complex 2 contains Int bound to *attB* and P1/P2. There is a 1:1 ratio of at *attB* to *attP* in this complex indicating that it is Int forming an intermolecular bridge between P1/P2 and *attB*. This complex is non-productive and is never able to undergo recombination no matter which components are added to the reaction (Peña and Hatfull 2000).

The isolation and characterization of these complexes is an important step in understanding L5 Int mediated recombination. The characterization of these complexes has provided a basis for understanding Int's ability to bind *attP* and *attB*. However, it is

Figure 11. L5 Int Complex Formation

A representative band shift experiment is shown in the middle with possible complexes on either side. In the presence of *attP* and mIHF L5 Int forms intasome (in the middle). Intasome is composed of Int –L5 (dark blue) forming an intramolecular bridge with *attP* (black lines) between core (long yellow box on *attP*) and P4/P5 (short yellow boxes on *attP*). A dimer of Int could be catalyzing intramolecular bridging (left), or a tetramer of L5 Int could be forming the bridge (right). DNA bending is catalyzed by mIHF (red ovals) that binds between arm and core site on *attP*. When *attB* is added to the reaction, the same interactions are made in synaptic complex 1 (upper band). However, *attB* is also bound by L5 Int subunits making an intermolecular bridge with P1/P2 (top cartoons). This could be occurring through L5 Int subunits that are bound together (right) or through different sets of four L5 Int subunits (left). In the presence of *attB*, synaptic complex 2 is also formed Lower band. This complex contains L5 Int creating an intermolecular bridge between P1/P2 and *attB*. This bridging could be carried out be either a dimer (left), or tetramer (right) of Int.



impossible to determine whether these complexes are intermediates in the recombination reaction or non-productive artifacts of electrophoresis.

I.E. Specific Aims

The goals of this research project are to characterization the L5 Int *attP* interaction using the following methods:

1. Characterize the L5 Int/*attP* interaction by creating mutants *attP* substrates and observing the effect of these mutations on L5 Int recombination and complex formation.
2. Creating L5 Int mutants and characterizing their ability to perform site-specific recombination and complex with *attP* and mIHF.
3. Characterize other phage integration systems revealed by comparative genomics and compare these to known phage recombination systems.

II. MATERIALS AND METHODS

II.A Bacterial Strains

II.A.i *Escherichia coli*

E. coli strains XL1-Blue and DH5 were used for DNA propagation. Strain BL21pLysS was used for protein expression.

II.A.ii Mycobacterium

Mycobacterium smegmatis mc^2155 is a high frequency transformation strain of *M. smegmatis* that was used for all bacterial experiments.

II.B Growth of bacterial cultures

E. coli strains were grown in LB broth (Difco) or on LB plates containing 1.5% agar at 37°C. Antibiotics for selection in *E. coli* were added to the following concentrations: kanamycin (kan), 20 g/ml; carbenicillin (Cb), 50 g/ml; tetracylcine (tet), 10 g/ml for liquid and 6.25 g/ml for plates; and hygromycin (hyg) 100 g/ml. *M. smegmatis* was grown in Middlebrook 7H9 broth or on Middlebrook 7H10 agar plates (Difco). Media for mycobacterial growth was supplemented with 10% ADC, 1 mM $CaCl_2$, 50 g/ml Cb and 10 g/ml chlorohexamide (chx) unless otherwise noted. Liquid media and plates used for replica plating, were supplemented with 0.05% - 0.1% Tween 80 to prevent clumping. Antibiotics for selection in mycobacteria were added to the following concentrations: kan, 20 g/ml; hyg, 100 g/ml; tet, 0.5 g/ml. For preparation of mycobacteria for phage infections, a saturated culture of *M. smegmatis* was diluted 1/100 and grown overnight in a baffled flask in medium lacking Tween 80.

Plating top agar was made using 7H9 medium with 0.7% agar and supplemented with 1% Glucose and 1 mM $CaCl_2$.

II.C Recombinant plasmids and cosmids

II.C.i Recombinant plasmids constructed by others

II.C.i.a pCP□R13

Plasmid pCP□R13 is derived from pMH94 (see below) by shortening the *attP* region to remove the P6/P7 arm-type sites leaving an P6/P7 minus *attP* that can be isolated by EcoRI/BamHI digestion. It is CB and Kan resistant, and it replicates extrachromosomally in *E. coli* and integrates in mycobacteria.

II.C.i.b pGH565

pGH565 has the entire L5 *int* gene inserted into the NdeI/HindII sites in the polylinker of pET21a (Novagen). Expression of the int gene is under the control of the T7 promoter and is inducible with IPTG.

II.C.i.c. pJL28 and pJL29

pJL28 and pJL29 were constructed by means of a three way ligation using an Xba I - Bam HI fragment from pGS74 containing 10 copies of the gp71 binding site, an XbaI - Bam H segment from pMD159 or pMD162 (for pJL28 or pJL29, respectively) and the vector pJL22 cut with Xba I.II. These plasmids contain *attL* and *attR* that can be isolated by EcoRI/BamHI digestion.

II.C.i.d. pMH plasmids

Plasmid pMH94 contains the *attP* and *int* region of L5 in a kanR and CbR backbone. A 2.1 kb piece of DNA in pMH94 has *attP* and the intact gene, *int.* This plasmid is able to insert into the *M. smegmatis* genome, but was primarily used in in vitro recombination assays.

II.C.i.e. pMK plasmids

pMK4 was created by cutting pCP32 at an XhoI site between core and P2, filling in the overhangs and ligating the blunt ends adding a 4 base pair insertion.
pMK6, 12, 13 and 17 were created using site-directed mutagenesis of pGL1 adding 11, 9, 6, and 13 base pairs, respectively, as well as a NcoI site.

II.C.i.f. pPSC1

pPSC1 is a pMH94 derived plasmid that has a single base pair mutation (G – A at position 1498) creating a NheI site between the N-terminal and middle domains of L5 Int. The mutation does not affect the amino acid composition of the protein.

II.C.ii. Recombinant plasmids constructed by Curtis Wadsworth

II.C.ii.a. pCW3

pCW2 was produced by performing site-directed mutagenesis on pCP□R13 at base pair 1296 (A – T) creating a BstBI site to the right of *attP* core. This created a NdeI/BstBI cassette containing the *attP* core. Site directed mutagenesis was carried out using the Stratagene site-directed mutagenesis kit with primers CW99-Cor1 and CW99-Cor2a, GTCGGCAGACACCCATTCTTCCAAACTAGCTACGCGGGTTCGATTCCCGTCGCC CGCTCCGC and GCGGAGCGGGCGACGGGAATCGAACCCGCGTAGCTAGTTTGG AAGAATGGGTGTCTGCCGAC, respectively.

II.C.ii.b. pCW21

pCW21 was created by exchanging the pMH94 Int for pPSC1 *int* gene by cutting both plasmids with XbaI and HinDIII and exchanging the 1955 base pair fragment of pMH94 for the same sized fragment from pPSC1.

II.C.ii.c. pCWCM plasmids

pCWCM plasmids were created by cutting pCW3 with NdeI and BstBI removing the *attP* core. This fragment was replaced with mutant core sequences derived from oligonucleotides that were engineered with NdeI and BstBI overhangs on the ends and annealed by boiling for 10 minutes and cooling to room temperature overnigh. The annealed oligonucleotides were kinased to add phosphates to the 5′ overhangs for efficient ligation. Oligonucleotides that were utilized are described in Table 1.

II.C.ii.d. pURL plasmids

pURL plasmids were produced by site directed mutagenesis that created a single base pair change in the *int* gene. This was accomplished by PCR mutagenesis using the Quick-Change Site-Directed Mutagenesis Kit (Stratagene). The oligonucleotides used in site directed mutagenesis are describe in Table 2.

II.D. Protein expression and purification

II.D.i. Integrase and Integrase Mutants

The L5 Int and mutant Int proteins were overexpressed by inducing *E. coli* strain BL21(DE3)pLysS containing pPSC1 or pURL plasmids grown at 30ºC with 1mM IPTG. Pelleted cells were frozen at -80º C then resuspended in TDE (20mM Tris pH 8, 1mM DTT, 1mM EDTA, 50mM NaCl) buffer containing 1mM phenylmethylsulfonyl flouride

Table 1. Creation of RBE Mutant attPs

Table 1. Creation of RBE Mutant *attP*s

Name	Mutation	Oligonucleotides
pCWCM1	R:ACCCATT□ CCACCTG L:ACTAGCT□ CCGATCG	TATGTGGTCGGCAGACCGAACCGCTTCCAACAGCTTGACGCGGGTT CGAACCCGCGTCAAGCTGTTGGAAGCGGTTCGGTCTGCCGACCACA
pCWCM2	R:ACCCATT□ CGATGCC L:ACTAGCT□ CAGGTGG	TATGTGGTCGGCAGACCCACCTGCTTCCAACCGATCGACGCGGGTT CGAACCCGCGTCGATCGGTTGGAAGCAGGTGGGTCTGCCGACCACA
pCWCML.5	R:No Change L:ACTAGCT□ CCGATCG	TATGTGGTCGGCAGACCCACCTGCTTCCAAACTAGCTACGCGGGTT CGAACCCGCGTAGCTAGTTTGGAAGCAGGTGGGTCTGCCGACCACA
pCWCMR.5	R:ACCCATT□ CCACCTG L:No Change	TATGTGGTCGGCAGACACCCATTCTTCCAACCGATCGACGCGGGTT CGAACCCGCGTCGATCGGTTGGAAGAATGGGTGTCTGCCGACCACA

Table 2. Creation of Middle Domain Int Mutants

Table 2. Creation of Middle Domain Int Mutants

Name	Mutation	Base Pair PPSC1	Amino Acid Change	Primers
pURL82	C → A	1461	W82L	GAGTACACCCGGAAGTTGCTCGTGGAGC GCTCCACGAGCAACTTCCGGGTGTACTC
pURL92	T → C	1432	T92I	CTCGCAGACGGCGCCAGGGATCTGTAC GTACAGATCCCTGGCGCCGTCTGCGAG
pURL106	C → G	1393	T106A	GAGCGCCGCATCTACGCGGTGCTAGGTG CACCTAGCACCGCGTAGATGCGGCGCTC
PURL113	T → A	1358	M113I	GGCGGTCACAGAGAAGACGCCAGCTCTG CAGAGCTGGCGTCTTCTCTGTGACCGCC
PURL115	G → C	1353	P115A	CACAGAGATGTCGGCAGCTCTGGTGCG CGCACCAGAGCTGCCGTCATCTCTGTG
PURL138	A → T	1295	H138L	CTGCCCGCCGGCTTGCCTACAACGTC GACGTTGTAGGCAAGCCGGCGGGCAG
PURL150	G → T	1256	A150E	GGTGATGAACACAGAGGTCGAGGACAAGCTG CAGCTTGTCCTCGACCTCTGTGTTCATCACC
PURL159	A → C	1227	N159T	CTGATCGCAGAGACCCCGTGCCGGATCG CGATCCGGCACGGGGTCTCTGCGATCAG
PURL160	G → C	1229	P160A	GATCGCAGAGAACGCGTGCCGGATCGAG CTCGATCCGGCACGCGTTCTCTGCGATC

(PMSF) and allowed to thaw on ice. PEI was added to lysates to 0.5% by volume, and DNA was allowed to precipitate on ice for 30 minutes before being pelleted by centrifugation. Proteins remaining in the lysate were precipitated on ice using 2.4 M ammonium sulfate. Proteins were pelleted by centrifugation. The pellet was redissolved in HMS. HMS buffer is 20 mM HEPES pH 8.0, 20 mM MES, 1 mM EDTA, 1 mM DTT and 50 mM NaAc. The proteins were then loaded onto a weak cation, carboxymethyl resin (CM) column using a Sprint BioCad chromatography system and eluted with a 50mM to 500mM NaCl gradient. Fractions eluting at approximately 200mM NaCl contained partially purified integrase. The protein containing fractions were determined by SDS-PAGE stained with Coomasie Brilliant Blue and combined. The combined fractions were then dialyzed in HMS containing 1.5 M ammonium sulfate overnight. The dialysate was then clarified by centrifugation and loaded onto a PE Column using the Sprint BioCad chromatography system. Bound protein was eluted between 1M and 500mM ammonium sulfate. The fractions containing Int were determined by SDS-PAGE, collected and dialyzed in TDE containing 20% glycerol overnight.

II.D.ii. mIHF

Mycobacterium smegmatis mIHF was overexpressed by inducing *E. coli* strain BL21(DE3)pLysS containing pMP30 plasmid grown at 37°C with 0.3 mM IPTG. Pelleted cells were frozen at -80° C then resuspended in TDE (20mM Tris pH 8, 1mM DTT, 1mM EDTA, 50mM NaCl) buffer containing 1mM phenylmethylsulfonyl flouride (PMSF) and allowed to thaw on ice. PEI was added to lysates to 0.5% by volume, and DNA was allowed to precipitate on ice for 30 minutes before being pelleted by centrifugation. Proteins remaining in the lysate were precipitated on ice using 50% ammonium sulfate.

Proteins were pelleted by centrifugation. The supernatant was collected, and a second ammonium sulfate precipitation was performed at 80% ammonium sulfate. The precipitated proteins were collected by centrifugation and the pellets were resuspended in HEM buffer (20 mM HEPES pH 8.0, 20 mM MES, 1 mM EDTA, 1 mM DTT). The proteins were then loaded onto a weak cation, carboxymethyl resin (CM) column using a Sprint BioCad chromatography system and eluted with a 50mM to 500mM NaCl gradient. Fractions eluting at approximately 200mM NaCl contained partially purified integrase. The protein containing fractions were determined by 15% SDS-PAGE stained with Coomasie Brilliant Blue and combined. These proteins were then loaded onto a POROS 20 HS column using a Sprint BioCad chromatography system and eluted with 50mM to 500mM NaCl gradient. Fractions eluted at approximately 200mM NaCl. The mIHF containing fractions were determined using 15% SDS-PAGE and combined before being dialyzed overnight against TDE containing 20% glycerol.

II.E. *In vitro* Recombination Assays

In vitro recombination was carried out using 3.4 nM *attP* and 46.5 nM – 700 nM *attB* with 11.9 nM Int and 280nM mIHF unless otherwise noted. Standard reaction conditions were: 20 mM Tris pH 7.5, 25 mM NaCl, 10 mM spermidine, 1 mM DTT, 1 mM EDTA, 0.1 mg/ml BSA. Reactions were carried out for 2 hours at 37ºC before being heat killed at 65ºC for 20 minutes. Products were separated on a 10% agarose gel and visualized by ethidium bromide staining under UV light. Products were quantified using BioRad Gel Documentation system and QuantityOne software. Supercoiled substrates were added as is. Linear substrates were created by digesting pCP□R13 with XhoI.

II.F. Radiolabeling DNA fragments

Labelled DNA fragments used for native gel electrophoresis were generated by a Klenow polymerase fill-in of the restriction digestion overhang with g\-P^{32} dATP. Fragments containing *attP* were obtained by BamH I / EcoR I digestion of pCPΔR13. Labelled *attB* containing DNA fragments were generated by kinasing with T4 poly-nucleotide kinase (PNK) of blunt ended annealed 42 base pair oligonucleotides in the presence of a\-P^{32} dATP.

II.G. Native gel electrophoresis

The reaction for complex formation was typically carried out in a 10 μl reaction with 20 mM Tris pH 7.5, 25 mM NaCl, 10 mM spermidine, 1 mM DTT, 1 mM EDTA, 0.1 mg/ml BSA and 100 μg/ml salmon sperm DNA. Protein concentrations were 28 nM – 5.6 μM nM mIHF and 12 nM Int, and 93 nM – 1.4 μM *attB* and 500 cpm of radiolabelled DNA were used unless otherwise indicated. The reactions were incubated on ice for 20 minutes and then immediately loaded on a 1X TBE, 5% acrylamide gel at 4ºC; a 10% Ficol loading buffer was used.

II.H. Computer programs

BLAST-P (Altschul *et al.*, 1990) and PSI-BLAST (Altschul *et al.*, 1997) were run using the web interface on the NCBI web site (http://www.ncbi.nlm.nih.gov). The Mac executable applications, BLASTALL and FORMATDB, were obtained from the NCBI ftp site (ftp://ftp.ncbi.nlm.nih.gov/blast/executables/). CLUSTAL-X version 1.8 (Jeanmougin *et al.*, 1998) and a mac executable version of CLUSTAL-W was downloaded from the Bio-Web web site

(http://web.tiscalinet.it/biologia/clustalx.hqx). All Mac applications were run on an iMac with a 333 MHz PowerPC G3 processor, 196 MB of RAM, and using the MAC OS 9.0.4 operating system. Compute pI/MW (Bjellqvist *et al.*, 1993), (Wilkins *et al.*, 1999) was used on the expasy web server (http://www.expasy.ch). Meltingpoint was written in PERL for this analysis and run on the iMAC using MacPerl.

Genome analysis was completed using GeneMark Version 2.4 (available at http://opal.biology.gatech.edu/GeneMark/genemark24.cgi) and GlimmerM (available at http://www.tigr.org/software). Annotated genes were compiled and open reading frames were cataloged using DNA Master (Lawerence). Putative open reading frames were compared to known sequences using BLAST-X and BlastInspector (Xu/Lewis).

II.I. Sequence acquisition and analysis

Searches were done using BLAST-P and PSI-BLAST on the NCBI web site searching a either the GenBank nr database or the GenBank protein annotation database. Searches done using BLASTALL were done on miscellaneous groups of proteins that were formatted into databases using FORMATDB. Sequence alignments and tree determination were done using CLUSTAL X. Trees were drawn using DrawGram from the PHYLIP package. The isoelectric focusing point, pI, was calculated using Compute pI/MW.

III. CHARACTERIZATION OF INT CORE BINDING

III.A. Introduction

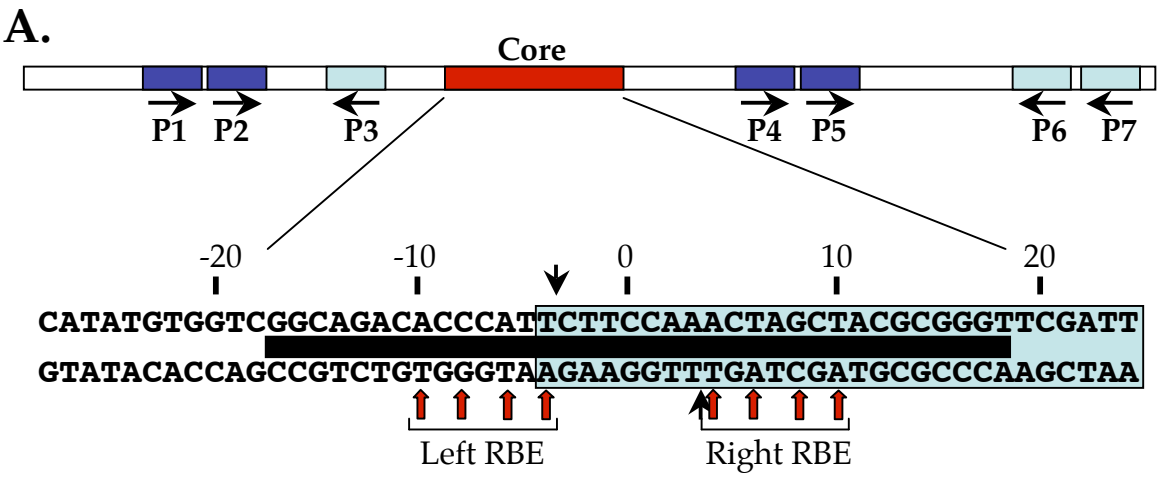
Lysogeny in Mycobacteriophage L5 involves the site-specific integration of the phage genome into the host chromosome (Lee and Hatfull 1993, Lee et al. 1991). The process of recombination is carried out by several molecules of a phage encoded tyrosine recombinase, integrase (Int), as well as a host encoded DNA binding protein (mIHF) (Lee and Hatfull 1993). These factors assemble into a complicated protein/DNA complex on the phage attachment site (*attP*). The 411 base pair L5-*attP* contains a core binding region composed of two 7 base pair recombinase binding elements (RBEs) flanking a 7 base pair overlap region where strand exchange occurs (Peña et al. 1996). There are 7 arm-type binding sites (P1 - P7) on either side of core that are arranged in pairs save P3. Of these, P1/P2 to the left of core and P4/P5 to the right of core are required for recombination (Figure 12 A) (Peña et al. 1997). Int interacts with *attP* heterobivalently, binding arm-type sites at its N-terminus and core at its C-terminus. Between the core and arm-type binding sites of *attP*, mIHF binds and bends *attP* thus delivering core to the C-terminal domain of Int (Peña and Hatfull 2000, Peña et al. 1999).

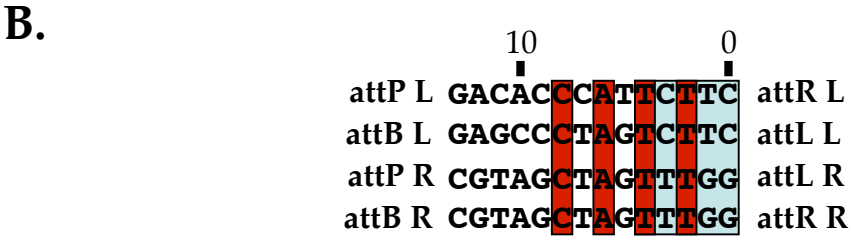
The formation of the heterobivalently bound Int/*attP* complex requires that Int bind core and arm-type binding sites differentially to keep the formation of unproductive complexes to a minimum. The RBEs of *attP* contain imperfect inverted repeats and *attP* and *attB* have different sequences their RBEs except for 4 base pairs involved in the inverted repeats (Figure 12 B) (Peña et al. 1996). Differences in the intervening base pairs may allow Int to bind *attP* core and *attB* diferentially, allowing Int to recognize the different substrates. Since *attL* core is identical in sequence to *attB*

Figure 12. The Elements of *attP*

A. *A linear representation of L5 attP.* P1 and P2, and P4 and P5 (dark blue) are arranged in pairs with a single base pair separating the sites and are utilized in recombination. P6 and P7 (light blue) have 2 base pairs separating the sites, and these sites along with P3 are not used in recombination. The core binding region (red) is between P3 and P4/P5. The sequence of the core binding region is shown below the linear representation of *attP*. The region of sequence homology with *attB* is shown in a light blue box. The thick black bar shows the location of Int binding as indicated by DNAse I footprinting. Small black arrows are used to show the sites of nicking in the 7 base pair overlap region. Bases of the RBEs that are involved in the inverted repeats are highlighted using red arrows.

B. *Sequence alignment of attP, attB, attR, and attL.* The sequences are labeled such that *attB* and *attP* half sites are on the left, and *attR* and *attL* half sites are on the right. The over lap region is in the light blue box. Conserved base pairs are in red boxes. It should be noted that the last base pair of the inverted repeat is absent in left RBE of *attB* and *attL*.


A.
Core
P1
P2
P3
P4
P5
P6
P7
-20
-10
0
10
20
CATATGTGGTCGGCAGACACCCATTCTTCCAAACTAGCTACGCGGGTTCGATT
GTATACACCAGCCGTCTGTGGGTAAGAAGGTTTGATCGATGCGCCCAAGCTAA
Left RBE
Right RBE


B.
10
0
attP L GACACCCATTCTTC attR L
attB L GAGCCCTAGTCTTC attL L
attP R CGTAGCTAGTTTGG attL R
attB R CGTAGCTAGTTTGG attR R

and *attR* is identical to *attP* core, differences in Int's binding to core based on RBE sequence could affect excision as well.

Int is thought to bind core after binding to arm-type sites at its N-terminus (Segall and Nash 1993). For this to be true, cross talk between domains must take place within Int protomers that are bound to arm-type sites. Conformational changes within these protomers could increase Int's affinity for core only on protomers that are bound to arm-type sites. Conversely, core binding could stimulate arm-type binding creating a two-way line of communication between the C and N terminal domains of Int.

III.B. Specific Aims

1. Utilize in vitro recombination assays to determine if Int is able to perform *attP*/ *attB*, *attP*/*attP*, *attP*/*attL*, and *attP*/*attR* recombination, thereby discerning the different substrates based on core binding.
2. Create mutants in RBEs, to determine if loss of inverted repeats flanking the 7 base pair overlap affect core binding and in vitro recombination.

III.C. Recombination Can Occur Between *attP* Cores

The RBE of *attP* core has a different sequence than the RBE of *attB* on the left side of the 7 base pair overlap. To determine whether this difference in RBE sequence affects Int's ability to distinguish between *attP* core and *attB*, *attP*/*attP* core recombination was carried out *in vitro*. *In vitro* reactions were set up under standard conditions containing 11.9 nM Int, 280 nM mIHF, 3.4 nM supercoiled *attP*-containing pMH94, and increasing concentrations of 42 base pair *attP* core or 42 base pair *attB* (9.3 – 930 nM). Reactions were incubated for 2 hours at room temperature before being stopped by heating to 65°C for 20 minutes. Recombination products were separated by electrophoresis on a

1% agarose gel and visualized by ethidium bromide staining under UV light (Figure 13).

Int is able to catalyze recombination at all concentrations of *attP* core tested, indicating that *attP*/*attP* core recombination can be accomplished even at relatively low concentrations of the 42 base pair *attP* core. Comparisons with *in vitro* reactions carried out with equimolar concentrations of *attB* suggest that recombination occurs with approximately equal efficiency with either substrate, indicating that the change in RBE does not affect the Int's recognition of core.

The efficiency of *attP*/*attP* core recombination was examined more closely by performing *in vitro* time course experiments. *In vitro* recombination was accomplished as described above in the presence of 270 nM 42 base pair core or 42 base pair *attB*. Recombination was allowed to occur for distinct amounts of time (15 – 120 minutes) before being stopped by heating to 65°C for 20 minutes. Products were separated on a 1% agarose gel and visualized by ethidium bromide staining under UV light (Figure 14).

The intensity of the product band observed in *attB*-containing reactions is greater than for *attP* core-containing reactions carried out for equal amounts of time, suggesting that Int catalyzes *attP*/*attB* recombination more efficiently than *attP*/*attP* core recombination. This shows that Int may be able to differentiate between *attP* core and *attB* core, but the difference in efficiency is slight and cannot be determined based on a single experiment.

L5 Int is unable to catalyze recombination at perceptible levels between *attP*'s containing arm-type binding sites. Presumably, Int protomers bind different molecules of *attP* forming a complex made up of Int protomers forming intramolecular bridges

Figure 13. *attP*/*attP* Core Recombination *in vitro*

In vitro recombination was used to test the ability of Int to perform recombination between *attP* and *attP* core. Reactions containing 11.9 nM Int, 3.4 nM pCPDR13, 280 nM mIHF, and increasing concentrations of *attB* or 42 base pair *attP* core (9.3 – 930 nM) were incubated at room temperature for 2 hours. Recombination was stopped by heating to 65ºC for 20 minutes, and products were separated by electrophoresis on a 1% agarose gel. Gels were stained with ethidium bromide, and visualized under UV light. Recombination on supercoiled plasmid yields linear product that runs between the lower supercoiled band and the higher nicked circular band. The amount of product present in the reactions increases as *attB* or 42 base pair *attP* core are added. These data show that Int is able to catalyze *attP*/core *attP* recombination at all concentrations of *attP* core tested with similar efficiency as *attP*/*attB* recombination.

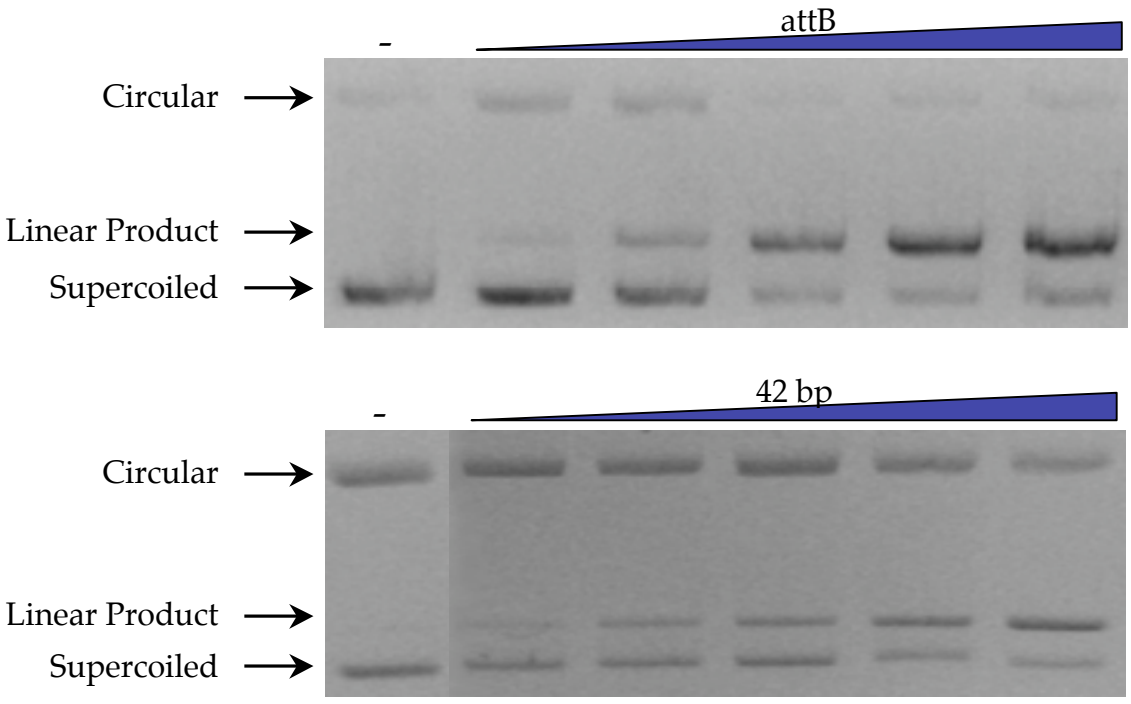
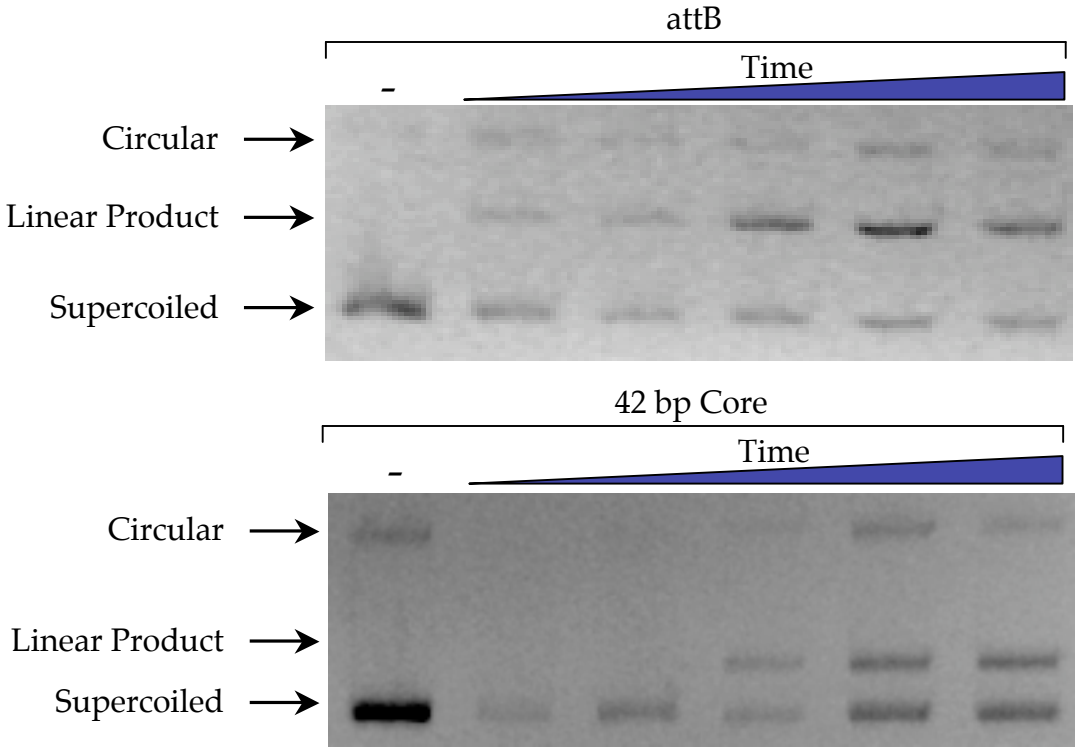

attB
-
Circular
Linear Product
Supercoiled
42 bp
-
Circular
Linear Product
Supercoiled

Figure 14. Time Course of *attP*/*attP* Core Recombination *in vitro*

In vitro recombination reactions containing 11.9 nM Int, 3.4 nM pCPDR13, 280 nM mIHF, and 70 nM *attB* or 42 base pair *attP* core were incubated at room temperature for discrete amounts of time (15 – 120 minutes) before reactions were heat killed and separated on a 1% agarose gel. Products were stained using ethidium bromide and visualized under UV light. As recombination is allowed to continue for longer periods of time, more linear product is observed running between faster moving supercoiled substrate and slower moving nicked circular DNA. These results suggest that Int is able to catalyze recombination slightly more efficiently during *attP*/*attB* recombination then *attP*/*attP* core recombination.


attB
Time
-
Circular
Linear Product
Supercoiled
42 bp Core
Time
-
Circular
Linear Product
Supercoiled

between P4/P5 and core (intasome), as observed in band shift assays performed in the absence of *attB*. Intasome formation inhibits the ability of Int to form the higher order recombinagenic complexes since both *attP*s would be competing for binding to arm-type sites. Differences in the region flanking the 7 base pair overlap region (RBEs) may also have an affect on binding, but *attP*/*attP* core *in vitro* recombination data suggests that this affect is secondary.

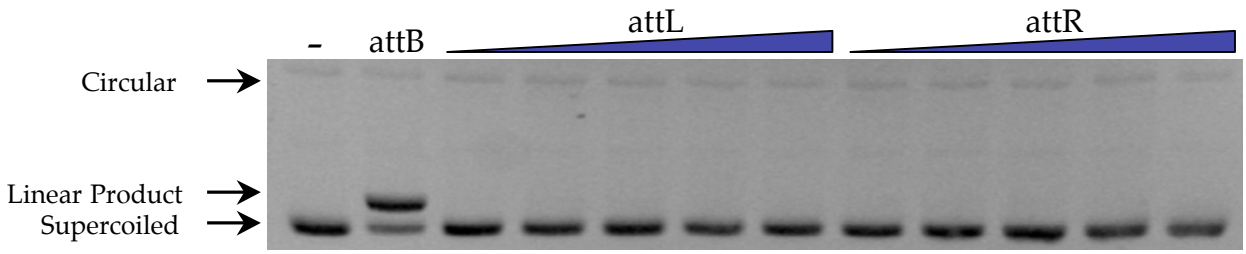
III.D. Int Cannot Catalyze *attP*/*attL* or *attP*/*attR* Recombination

Excision of the L5 lysogens involves a recombination reaction between *attL* and *attR*, and can only take place in the presence of a phage-encoded excisionase (xis). Xis mediates *attL*/*attR* recombination by stabilizing the Int/*attR* complex (Lewis and Hatfull, 2000). Recombination between *attP*/*attR* and *attP*/*attL* does not appear to occur *in vivo* since tandem insertions have never been observed. The right side RBE and the left side RBE of *attB* with P4/P5 make up *attL*, and *attR* consists of the left side RBE of *attP* and the right of *attB* (Figure 12 B). If Int binds *attL* and *attR* cores differently then *attP*, the inability of *attP* containing intasome to recognize *attL* or *attR* RBEs could block recombination.

To test this hypothesis, 342 base pair *attL* or *attR* containing fragments of pJL38 or 39, respectively, were used as substrates in place of *attB* in reactions containing 11.9 nM Int, 280 nM mIHF, 3.4 nM supercoiled *attP* containing pCP□R13, and increasing concentrations of *attL* or *attR*. Reactions were carried out for 2 hours at room temperature and stopped by heat killing at 65ºC for 20 minutes. Recombination products were separated on a 1% agarose gel and visualized by ethidium bromide staining under UV light (Figure 15).

Figure 15. *attP*/*attL* and *attP*/*attR* Recombination *in vitro*

Reactions containing 11.9 nM Int, 280 nM mIHF, 3.4 nM supercoiled *attP* containing pCP□R13, and increasing concentrations of an *attL* or *attR* containing fragment of pJL38 or 39, respectively, were allowed to perform recombination for 2 hours at room temperature. Reactions were stopped by heating to 65°C for 20 minutes, separated on a 1% agarose gel, and visualized by ethidium bromide staining under UV light. Linear product is seen in *attB* containing control reactions running between the faster moving supercoiled substrate band and the slower moving nicked circular band. No product is seen at any concentration of *attL* or *attR* showing that Int is unable to catalyze recombination between *attP* and *attL* or *attR*.

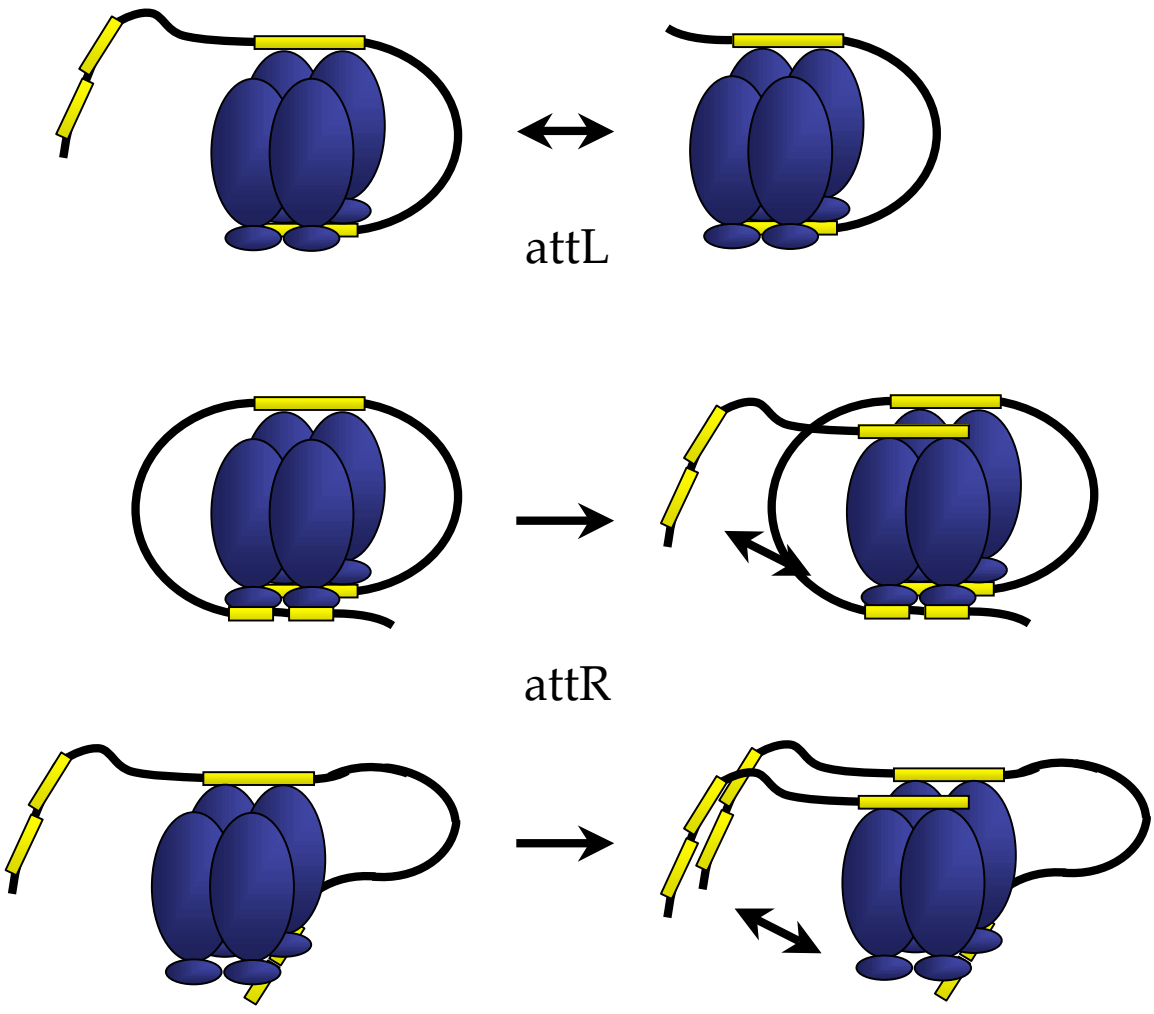

-
attB
attL
attR
Circular
Linear Product
Supercoiled

Int is unable to catalyze *attP*/*attL* or *attP*/*attR* recombination at any of the *attL* or *attR* concentrations tested. This implies that Int can recognize neither *attL* nor *attR* in the recombinagenic complex. This could show that Int is unable to recognize the *attL* or *attR* cores, indicating that at some level the recognition sequences, RBEs, on either side of the 7 base pair overlap region are important for the Int/core recognition. However, it seems more likely that the presence of arm-type binding sites is blocking recombination when *attL* and *attR* substrates are used in place of *attB*. There is competition arm-type binding of Int between *attP*, and *attL* and *attR*. This is likely the case for *attL*, since *attL* is known to form a stable Int/*attL* complex following recombination. P4/P5, which remains occupied after recombination, is part of *attL* (Pena, 1999). P1/P2 is not occuoied following recombination unless Xis is present, indicating that *attR* does not form a complex with Int (Lewis and Hatfull, 2000). Therefore, the blocking of recombination by a heterobivalently bound *attR* would seem unlikely. The reduced affinity of Int for *attR* arm-type sites would render *attR* less able to compete with *attP* for Int binding. Therefore, *attR* must be rejected by a different method. If binding to core stimulates arm-type site binding, *attR* and *attP* could be competing for binding to the same N-terminal domains when *attR* is brought into the recombinagenic complex. Presumably, binding of *attP*'s P1/P2 is more probable since it is already part of the Int/*attP* complex, so *attR* is effectively eliminated as substrate (Figure 16).

To determine whether Int is able to recognize *attR* core while bound to *attP*, electrophoretic mobility shift assays (EMSA) was used. If Int is able to recognize and bind *attR* core, an Int/*attR* complex should appear. Intasome will decrease as *attR* is added to the reactions if *attR* is blocking recombination by competing with *attP* for Int binding. These assays were performed under standard conditions with 11.9 nM Int, 560 nM mIHF, radio-labeled *attP* and increasing concentrations of *attL* or *attR*. Complexes

Figure 16. Models of *attL* and *attR* Recombination Inhibition

As indicated by *in vitro* recombination, Int is unable to perform *attP*/*attL* recombination. In EMSA competition assays, the ability of Int to form intasome is lost at high *attL* concentrations. This is probably due to Ints ability to form stable complex with *attP* and *attL*, and *attL* is competing with *attP* for Int binding. Int does not stably bind *attR*, and recombination could be blocked by a different method. If P1/P2 binds Int prior to *attB* capture, *attR* would not be able to join the integration complex because it would be competing with *attP* for binding at arm-type binding sites. The same would be true if *attB* capture occurs at the same time as *attP* binding as described in Figure 12. Competition for arm-type binding would block *attR* binding. No mIHF is shown in these diagrams to show the path of the DNA more clearly.


attL
attR

were formed on ice for 1 hour and were separated by electrophoresis on a 5% native poly-acrylamide gel. Results were visualized by autoradiography (Figure 17).

In the absence of *attB*, Int forms stable intasome. The density of the intasome band is lost as *attL* is titrated into the reaction indicating that *attP* is competing with *attL* Int binding. At high *attL* concentrations, very little intasome is still visible, and more *attP* is running as free substrate. No other complexes are formed showing that Int is binding both substrates simultaneously. This also appears to be the case for *attR*, although inhibition of intasome formation is not observed until higher *attR* concentrations reflecting the inability of *attR* to complex with Int in the absence of Xis. These data suggest that the Int/*attR* core interaction is not as stable as *attL*, but *attR* is able to compete for Int binding. Since no new complexes are formed, the Int/*attR* complex alluded to above is either not formed, or very unstable.

The presence of arm-type binding sites on recombination substrate is the determining factor in substrate choice by Int. The composition of the RBEs may affect the direction of *attB* binding, but does not appear to affect the ability of Int to perform recombination.

III.E. RBE Mutants Cannot Be Used As Substrate By L5 Int

To further examine the Int/core relationship, *attP*s were created that contain mutations in core RBEs. Two sets of mutations were created. The bases that form the inverted repeats in both RBEs are mutated in pCWCM1. Intervening bases are left as wild type. Every base in the 7 base pair inverted repeats in both RBEs are mutated in pCWCM2 (Figure 18 A).

In vitro recombination was carried out on pCWCM1 and pCWCM2 in reactions containing 11.9 nM Int, 280 nM mIHF, 70 nM *attB*, and ~3.4 nM supercoiled substrate,

Figure 17. EMSA of *attP*/*attL* or *attP*/*attR* Complexes

11.9 nM Int, 560 Nm mIHF, radio labeled *attP* and increasing concentrations of *attL* or *attR* were mixed under standard conditions and allowed to form complexes on ice for 1 hour. Complexes were then separated on a 5% poly-acrylamide gel and visualized by autoradiography. Intasome is formed in the absence of *attL*. As the concentration of *attL* is increased, the intensity of the intasome band decreases indicating that *attL* is competing with *attP* for Int binding. The same is true for *attR*, however the inhibition of Intasome formation is only seen at high *attR* concentrations. These data suggest that *attL* and *attR* are competing with *attP* for binding to *attP* and inhibiting Ints ability to perform recombination.

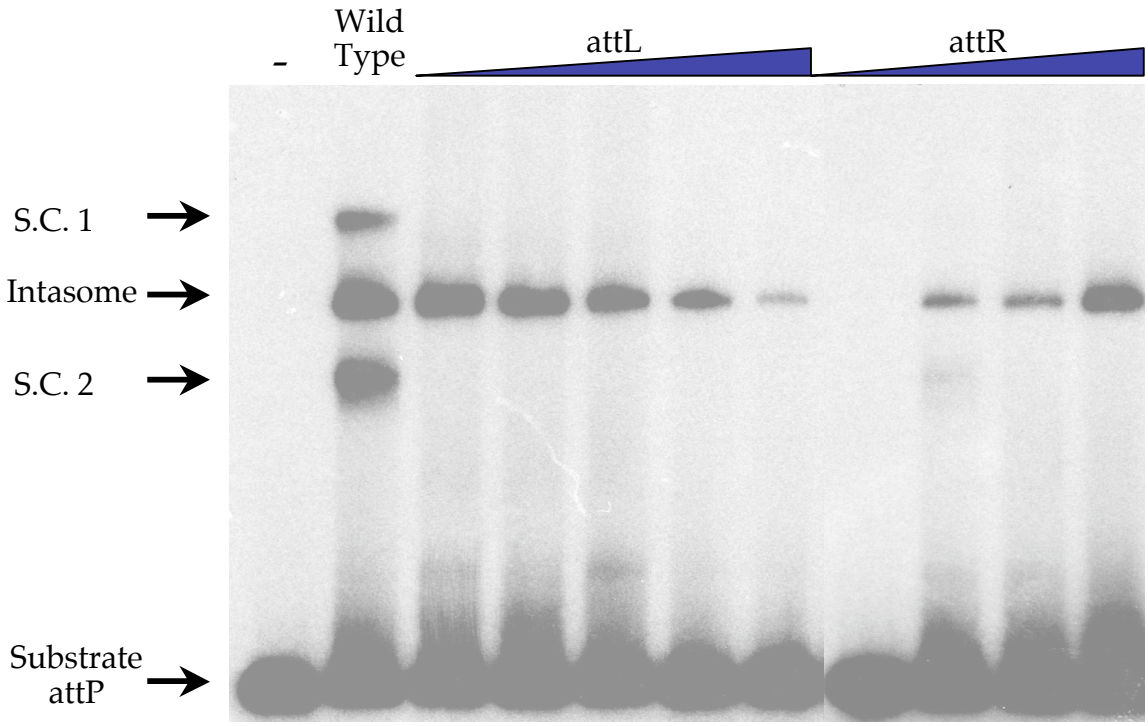
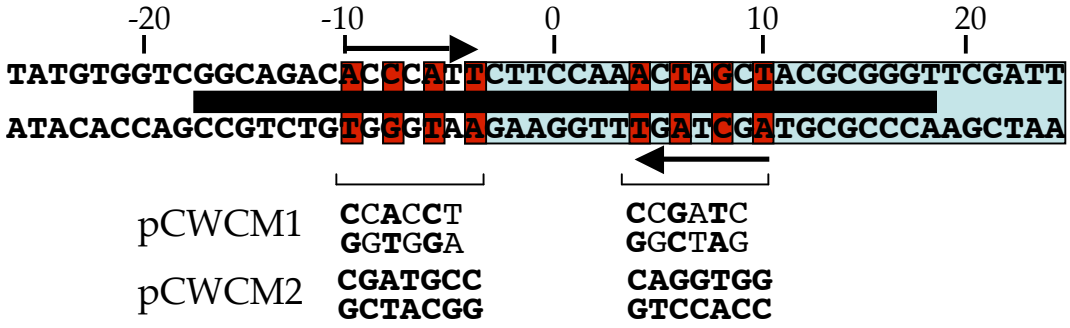

-
Wild Type
attL
attR
S.C. 1
Intasome
S.C. 2
Substrate attP

Figure 18. pCWCW1 and pCWCM2 *attP* Core Mutants

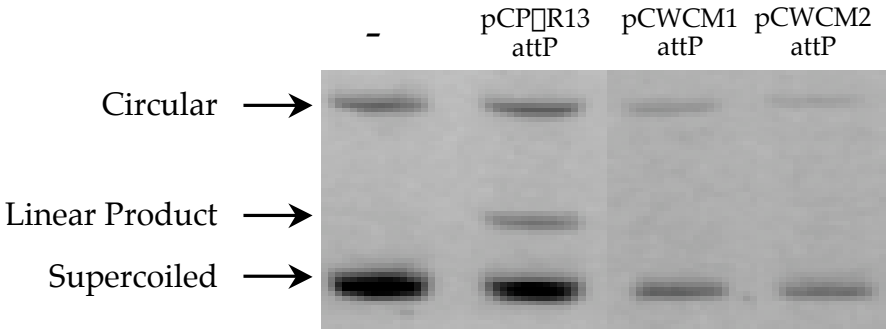
A. The sequence of *attP* core is shown here. Sequence shared with *attB* is shown in the light blue box. The seven base pair overlap region is flanked on either side imperfect inverted repeats that make up *attP*s RBEs. RBEs are highlighted with black arrows, and repeated base pairs are in red boxes. Mutations made to RBEs are shown below as indicated with changed base pairs in bold. pCWCM1 is shown above pCWCM2.

B. *In vitro* recombination was carried out using supercoiled pCWCM1 and pCWCM2 as substrate in reactions containing 11.9 nM Int, 280 nM mIHF, and 70 nM *attP* under standard conditions. Recombination took place for 2 hours at room temperature before being stopped by heating to 65°C for 20 minutes and separated on a 1% agarose gel. Products were stained using ethidium bromide and visualized under UV light. Linear product is observed in the control reaction using wild type *attP* containing pCP□R13 between the lower supercoiled substrate band and higher nicked supercoiled bands. No product is observed in reactions containing pCWCM1 or pCWCM2. These results show that Int is unable to catalyze recombination when pCWCM1 or pCWCM2 are used as substrate.

A.


-20
-10
0
10
20
TATGTGGTCGGCAGACACCCATTCTTCCAAACTAGCTACGCGGGTTCGATT
ATACACCAGCCGTCTGTGGGTAAGAAGGTTTGATCGATGCGCCCAAGCTAA
pCWCM1
CCACCT
GGTGGA
CCGATC
GGCTAG
pCWCM2
CGATGCC
GCTACGG
CAGGTGG
GTCCACC

B.


-
pCP□R13 attP
pCWCM1 attP
pCWCM2 attP
Circular
Linear Product
Supercoiled

pCP□R13, pCWCM2 and pCWCM1. Reactions were incubated for 2 hours at room temperature and stopped by heating to 65°C for 20 minutes. Products were separated on a 1% agarose gel and visualized by ethidium bromide staining (Figure 18 B).

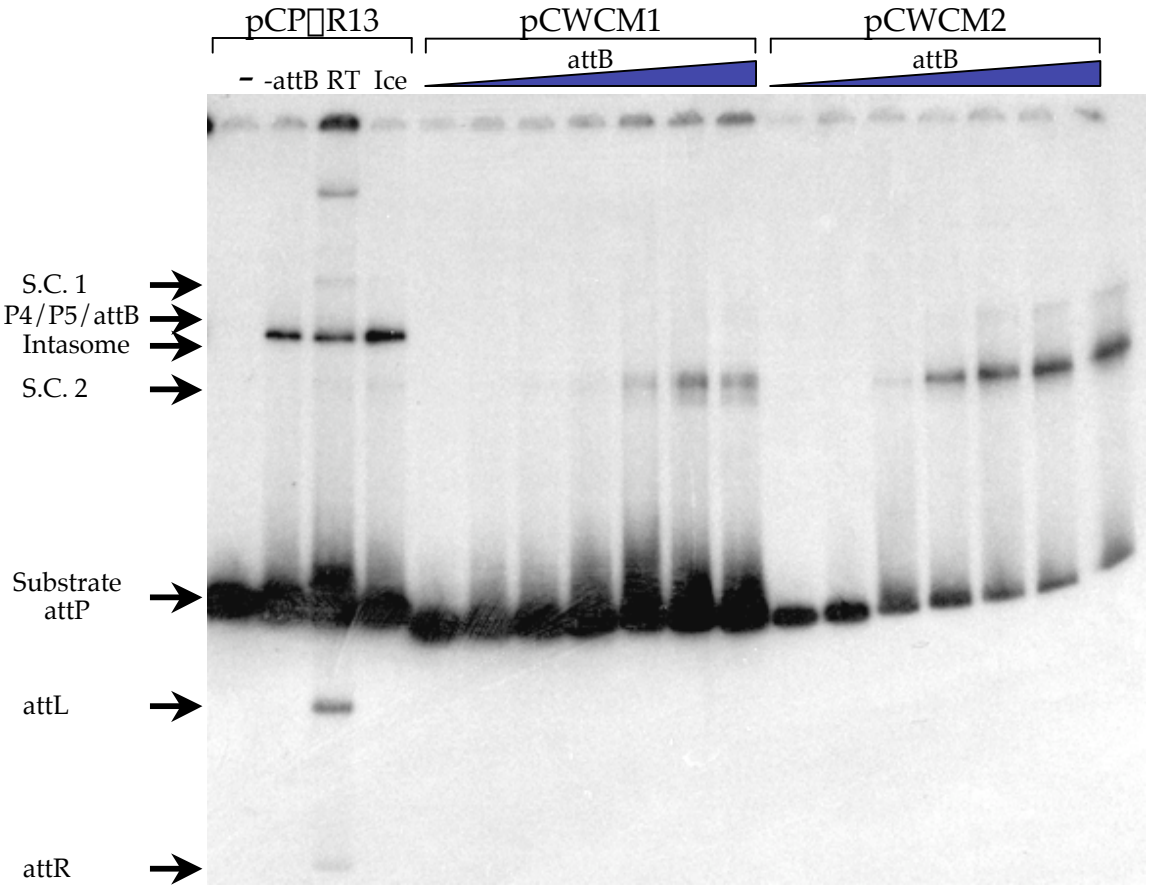
L5 Int can use neither core mutant as substrate for recombination indicating that Int is unable to recognize the 7 base pair overlap of core without the flanking RBEs. Since no product is observed in reactions using pCWCM1 as substrate, Int must contact the conserved base pairs of the inverted repeat and use these to recognize the core. The non-inverted repeat base pairs may contact Int, but are unable to catalyze core recognition alone.

The inability of Int to perform recombination using pCWCM1 and pCWCM2 suggest that Int is unable to recognize the mutant *attP* cores. To test the ability of Int to bind mutant *attP*s, band shift analysis was used to observe the binding of Int to mutant *attP*s directly. Int makes several complexes when mixed with mIHF, *attB*, and radio-labeled *attP* that can be separated by native PAGE. Of these, intasome and synaptic complex 2 require core binding for their formation. Therefore, the ability of Int to recognize pCWCM1 and pCWCM2 *attP* cores can be examined by monitoring the formation of these complexes. Band shifts were carried out in the presence of 11.9 nM Int, 560 nM mIHF, 140 nM *attB*, and radio labeled *attP*, a 342 base pair BamHI/EcoRI fragment of pCPDR13, pCWCM1, or pCWCM2 containing their respective *attP*s. Complexes were allowed to form on ice for 20 minutes before being separated on a 5% non-denaturing poly-acrylamide gel run at 4°C and were visualized by autoradiography (Figure 19).

Int forms 3 distinct complexes in the presence of wild type *attP* when incubated on ice. Intasome is composed of Int forming an intramolecular bridge between P4/P5 and core. In synaptic complex 1 that runs above intasome, Int is forming the same

Figure 19. pCWCM1 or pCWCM2 *attP* Complex Formation

EMSA was carried out using 11.9 nM Int, 560 nM mIHF, 140 nM *attB*, and radio-labeled 342 base pair BamHI/EcoRI fragment of pCPDR13, pCWCM1, or pCWCM2 *attP*s was used to observe Int *attP*/core binding directly. Complexes were allowed to form on ice for 1 hour before being separated on a 5% non-denaturing poly-acrylamide gel, and visualized by autoradiography. In the absence of *attB*, Int binds wild type *attP* forming intasome, and synaptic complexes 1 and 2 at low levels when *attB* is added. Int forms only synaptic complex 2 in the presence of pCWCM1 and pCWCM2 *attP*. In pCWCM1 *attP* containing reactions, synaptic complex 2 is formed weakly, where as Int forms synaptic complex 2 and P4/P5/*attB* containing complexes well in the presence of pCWCM2 *attP*


pCP□R13
pCWCM1
pCWCM2
-
-attB
RT
Ice
attB
attB
S.C. 1
P4/P5/attB
Intasome
S.C. 2
Substrate
attP
attL
attR

intermolecular bridge as intasome as well as an intermolecular bridge between P1/P2 and *attB*. Synaptic complex 2 runs below intasome and contains *attB* as well. However, core is not bound in this complex, so synaptic complex 2 probably represents Int forming an intermolecular bridge between P1/P2 and *attB*. A fourth complex that represents an Int mediated intermolecular bridge between P4/P5 and core is observed at high *attB* concentrations and runs just above intasome.

Int cannot form intasome or synaptic complex 1 when either pCWCM1 or pCWCM2 *attP*s are used as substrate in EMSA. Synaptic complex 2 is formed in the presence of both pCWCM1 and pCWCM2 *attP*s indicating that Int is still able to recognize the arm type binding sites of these substrates. Reactions containing pCWCM2 *attP* form complexes that appear very similar to reactions containing wild type *attP* in the absence of mIHF indicating that Int is not recognizing core. Interestingly, pCWCM2 *attP* appears to be used as substrate for Int bridging complexes more readily then pCWCM1 *attP* as evidenced by stronger synaptic complex 2 formation at lower concentrations of *attB* and formation of P4/P5/*attB* intermolecular complexes at high *attB* concentrations.

Weak synaptic complex 2 formation by pCWCM1 *attP* could indicate that weak binding is occurring between Int and the mutant core. Transient Int/*attP* core interactions would allow the formation of unstable intasome in the presence of pCWCM1 *attP* that run as smears around the substrate band. This interaction is able to out compete *attB* for binding to Int at low *attB* concentrations, and no synaptic complex 2 is observed. At higher *attB* concentrations, *attB* is able to out compete the reduced affinity *attP* core for binding allowing weak formation of synaptic complex 2.

III.F. *attP* Mutants Containing Mutations in to the Left of the 7 Base Pair Overlap Can Support Recombination but Not to the Right

Core mutants with only one RBE mutated while the other remains wild type were created to test Ints ability to perform recombination when only half of the inverted repeat is provided. Like pCWCM1, only the base pairs in the inverted repeats were mutated in each half site. The RBE to the left of the 7 base pair overlap is mutated in pCWCML.5, while pCWCMR.5 contain mutations in the right half site (Figure 20 A).

In vitro recombination assays containing 11.9 nM Int, 560 nM mIHF, 140 nM *attB*, and ~4 nM supercoiled pCWCML.5 or pCWCMR.5 were used to test the ability of these mutants to be used as substrate in L5 Int recombination. Reactions were carried out at room temperature for 2 hours and stopped by heating to 65°C for 20 minutes. The products were then separated by electrophoresis on a 1% agarose gel and visualized by ethidium bromide staining under UV light (Figure 20 B).

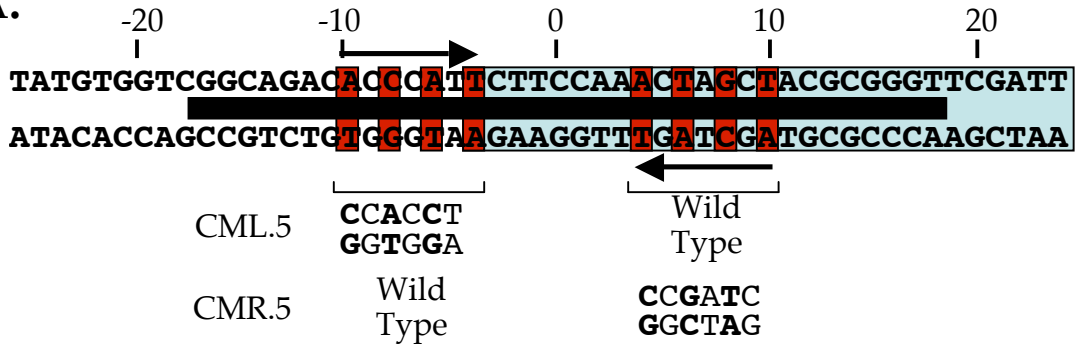
Low levels recombination occurs when pCWCML.5 is used as substrate for *in vitro* recombination assays. However, pCWCMR.5 does not support recombination under the same conditions. The mutation in the pCWCMR.5 is on the same side of *attP* as arm-type binding sites P4/P5, and P4/P5 is occupied in intasome making it a good candidate for being the partner of *attP* core during intermolecular bridging. Int is probably binding to P4/P5 and then wrapping *attP* around the Int complex with the help of mIHF. Therefore, *attP* core is delivered to the C-terminal domain of Int. In this case, the RBE to the right of core is probably contacting Int first. Since pCWCMR.5 *attP* is mutated at the first core contact point, the core cannot be bound because this interaction is not made. This interaction is made by pCWCML.5, so the core can be

Figure 20. Construction and Characterization of pCWCML.5 and pCWCMR.5

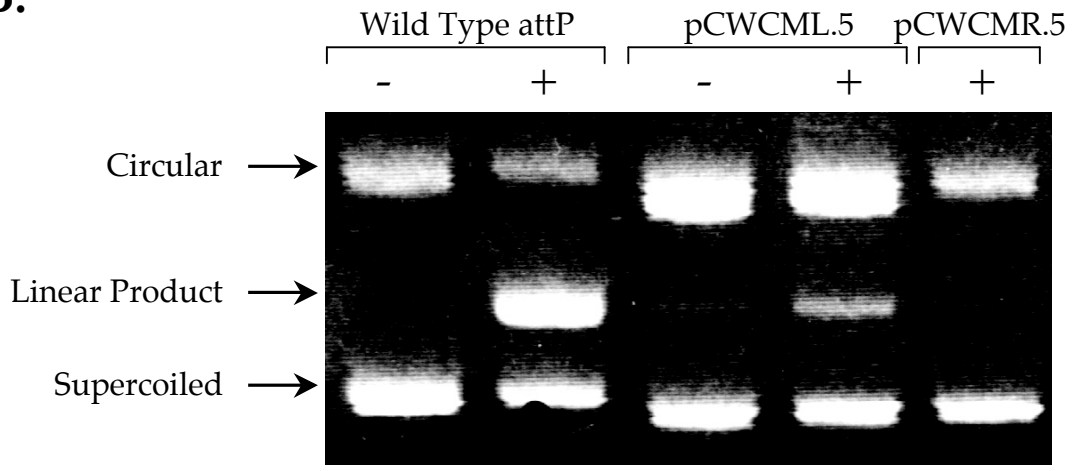
A. The *attP* core DNA sequence is shown here with a light blue box around the sequence shared with *attB*. Inverted repeats within RBEs flanking the 7 base pair overlap region (bracketed with black arrows indicating the direction of the inverted repeats) are shown in red boxes. The base pairs changed are shown in bold below the bracketes non-bolded base pairs were unchanged. Wild type indicates that none of the base pairs were mutated. Plasmid names are given to the left.

B. *In vitro* recombination was carried out using supercoiled pCWCML.5, pCWCMR.5, and wild type *attP* containing pCP□R13 as primary substrate with 11.9 nM Int, 280 nM mIHF and 70 nM *attB* under standard conditions. Recombination was allowed to occur for 2 hours at room temperature, reactions were heat killed then separated on a 1% agarose gel, and visualized by ethiduim bromide staining under UV light. Linear product is formed in the presence of wild type *attP* as well as pCWCML.5 but not when pCWCMR.5 is used as supercoiled substrate as observed by a product band between faster moving supercoiled and slower moving nicked circular plasmid. These data show that pCWCML.5 is able to carry out limited recombination while pCWCMR.5 was unable to support recombination.

A.


-20
-10
0
10
20
TATGTGGTCGGCAGACACCCATTCTTCCAAACTAGCTACGCGGGTTCGATT
ATACACCAGCCGTCTGTGGGTAAGAAGGTTTGATCGATGCGCCCAAGCTAA
CML.5
CCACCT
GGTGGA
Wild
Type
CMR.5
Wild
Type
CCGATC
GGCTAG

B.


Wild Type attP
pCWCML.5
pCWCMR.5
-
+
-
+
+
Circular
Linear Product
Supercoiled

bound initially even though the affinity for pCWCML.5 *attP* is lower then wild type. This allows recombination to occur at low levels when pCWCML.5 is used as substrate.

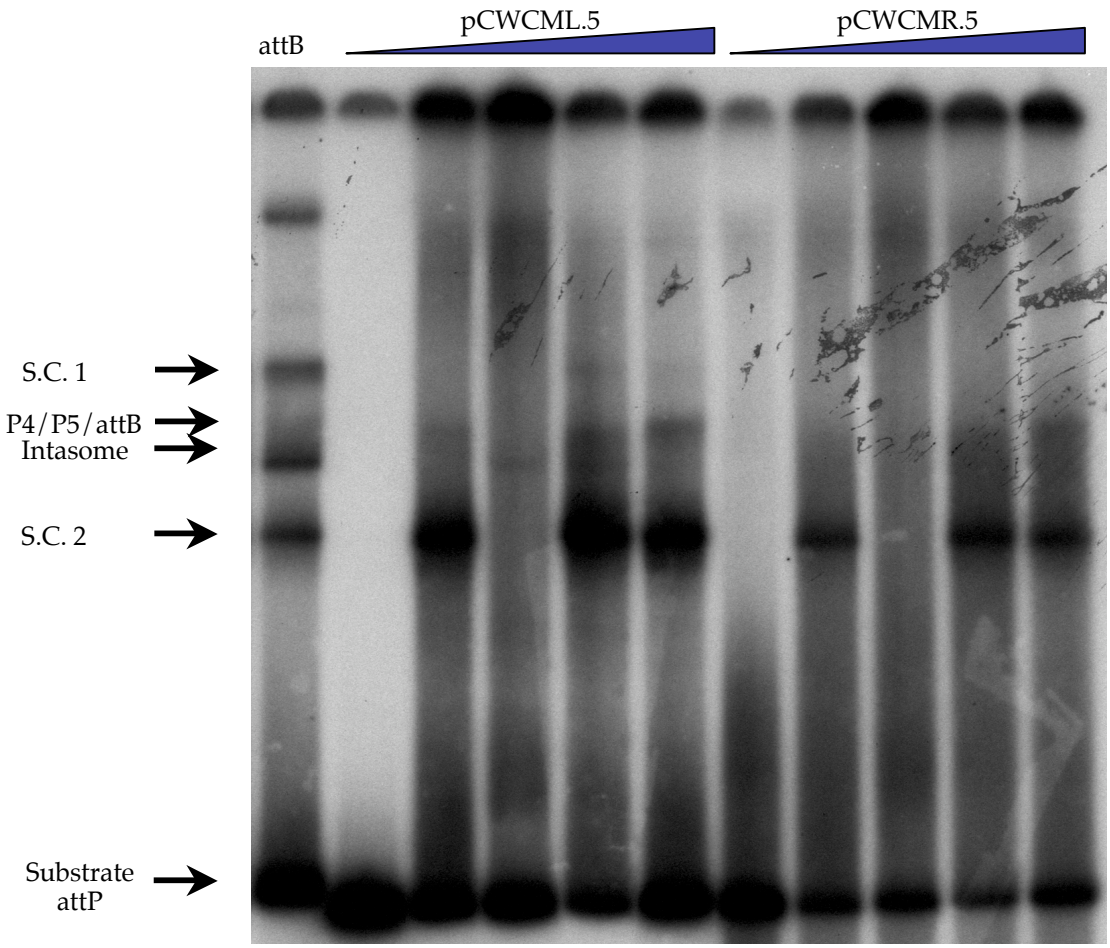
An interesting perspective on the Int/core interaction can be gained from the study of pCWCML.5 and pCWCMR.5. To further characterize these mutants, EMSA was performed. A 342 base pair *attP*-containing fragment of pCWCML.5 and pCWCMR.5 was cut from the full-length plasmid and radio-labeled. These substrates were used in EMSA reactions containing 11.9 nM Int, 560 nM mIHF, and 9.3 – 930 nM *attB*. Complexes were formed on ice for 1 hour and separated by electrophoresis on a 5% non-denaturing poly-acrylamide gel. Results were visualized by autoradiography (Figure 21).

Very small amounts of intasome and synaptic complex 1 that is visible at high concentrations of *attB* are formed when pCWCML.5 *attP* is used as substrate. Synaptic complex 2 is highly favored in reactions containing pCWCML.5 and can be seen in every lane containing *attB*. Intasome or synaptic complex 1 are not formed when pCWCMR.5 *attP* ise used under the same conditions, but synaptic complex 2 is observed at all *attB* concentrations and the P4/P5/*attB* complex is formed at high concentrations of *attB*.

The formation of small amounts of intasome and synaptic complex 1 with pCWCML.5 *attP*, but not pCWCR.5 *attP* illustrate the importance of RBE binding in the formation of these complexes. The RBE to the right of core must form a stable bond with Int in order for the complex to form properly. Even if the RBE to the left of the overlap can form wild type interactions with core, intasome cannot form, and recombination cannot take place.

Figure 21. pCWCML.5 and pCWCMR.5 *attP* Complex Formation

EMSA was performed using a radio labeled 342 base pair *attP* containing fragments of pCWCML.5, pCWCMR.5, or pCP□R13. These substrates were mixed with 11.9 nM Int, 560 nM mIHF and140 nM *attB* under standard conditions, and complexes were formed on ice for 1 hour before being separated on a 5% non-denaturing poly-acrylamide gel. Complexes containing radio-labeled *attP* were visualized by autoradiography. Int forms intasome as well as synaptic complexes 1 and 2 in the presence of *attB*. Synaptic complex 2 and P4/P5/*attB* are formed when pCWCMR.5 *attP* is used as radio-labeled substrate, and Int is able to form small amounts of synaptic complex 1 and intasome at high *attB* concentrations in reactions containing radio-labeled pCWCML.5 *attP*.


attB
pCWCML.5
pCWCMR.5
S.C. 1
P4/P5/attB
Intasome
S.C. 2
Substrate
attP

III.G. DNA Fragments Containing Arm-Type Binding Site Do Not Inhibit Recombination in Trans

Int is thought to bind *attP* by first binding to P4/P5 (Peña and Hatfull 2000). This activates core binding by the Int protomers contacting P4/P5. Since *attP* core is at higher concentration locally, Int binds *attP* core preferentially. This ensures that intasome is formed and not errant complexes containing *attB*. Int does not bind arm-type sites with great enough affinity that complexes containing Int and P4/P5 or P1/P2 containing DNA fragments can be separated by EMSA, but these fragments could bind Int with great enough affinity to block *attP* arm-type binding.

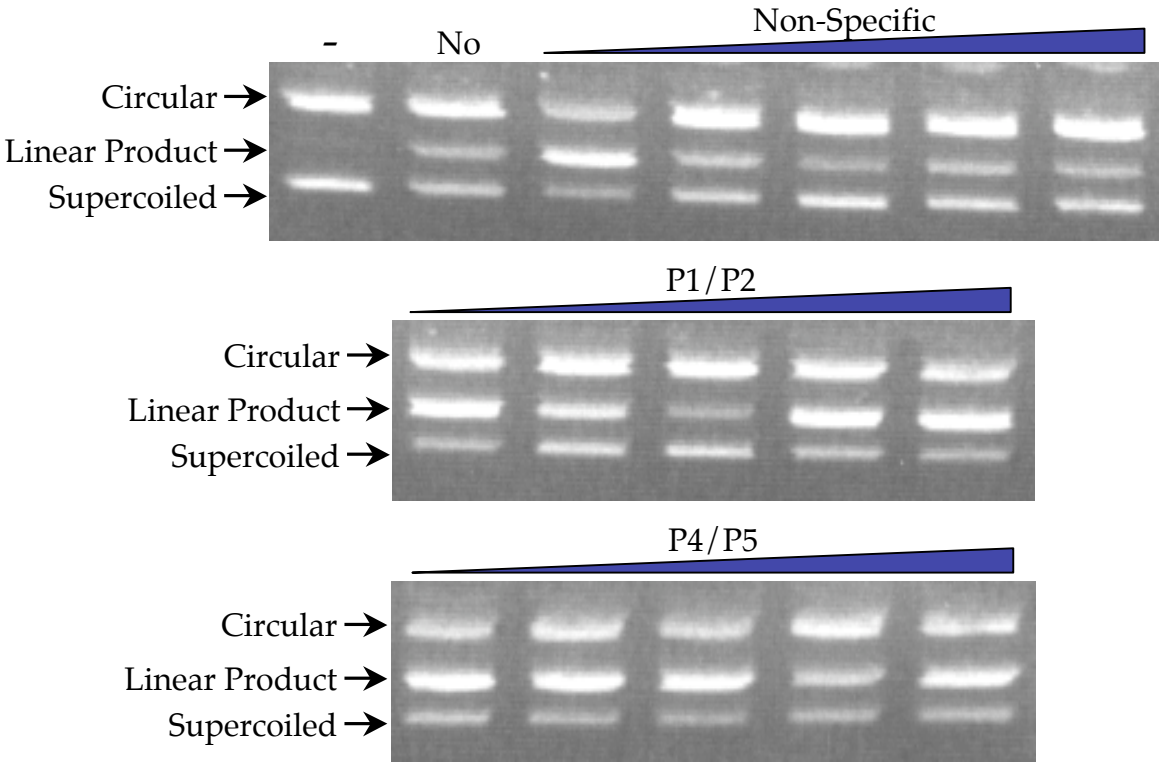
In vitro recombination assays in the presence of increasing concentrations of 43 base pair oligonucleotides containing arm-type sites was used to test this hypothesis. Reactions containing 11.9 nM Int, 280 nM mIHF, 70 nM *attB*, and increasing concentrations of non-specific DNA or P1/P2 or P4/P5 containing DNA fragments were allowed to undergo recombination at room temperature for 2 hours before being stopped by heating to 65ºC for 20 minutes. The products of recombination were then separated on a 1% agarose gel and stained using ethidium bromide. Results were observed under UV light (Figure 22).

To ensure that the non-specific substrate was unable to support recombination, *in vitro* recombination was carried out with non-specific substrate in the absence of *attB*. No product was observed (not shown). When the same non-specific substrate was titrated into reactions containing *attB*, product formation was not affected indicating that the 43 base pair oligonucleotide was not interfering with recombination.

Providing P4/P5 or P1/P2 containing oligonucleotides in trans does not affect the ability of Int to perform recombination. These results reflect either the inability of

Figure 22. *attP*/Arm Type Binding Site *in vitro* Competition

In vitro recombination was performed using 11.9 nM Int, 280 nM mIHF, 70 nM *attB*, and 4 nM supercoiled *attP* containing pCP□R13. 43 base pair P1/P2, P4/P5 or non-specific oligonucleotides was titrated into these reactions, and recombination was allowed to continue for 2 hours at room temperature before being stopped by heating to 65ºC for 20 minutes. Products were separated on a 1% agarose gel and visualized by ethidium bromide staining under UV light. Linear product running between supercoiled and linear substrate is observed in all *in vitro* reactions indicating that P1/P2 or P4/P5 oligonucleotides neither inhibit nor stimulate recombination.


-
No
Non-Specific
Circular
Linear Product
Supercoiled
P1/P2
Circular
Linear Product
Supercoiled
P4/P5
Circular
Linear Product
Supercoiled

arm-type sites to compete with full-length *attP* for binding to Int, or the inability of these oligonucleotides to remain in complex with Int following core binding. This indicates that binding of Int to both the arm and core-binding sites of *attP* requires the linker region between the binding sites. The linker region could tether *attP* to Int during in conformation changes in either the N or C-terminal domains that could cause the substrate to be released from the complex. Stabilization of Int/*attP* complex may involve mIHF since it binds within the linker region. The binding of Int to arm-type binding has been thought to stimulate Ints binding to core. The inability of arm-type site containing oligonucleotides to remain bound to arm-type sites suggest that cooperative interactions allow Int to remain bound to arm-type sites after core has been bound. This is indicative of two-way communication between domains.

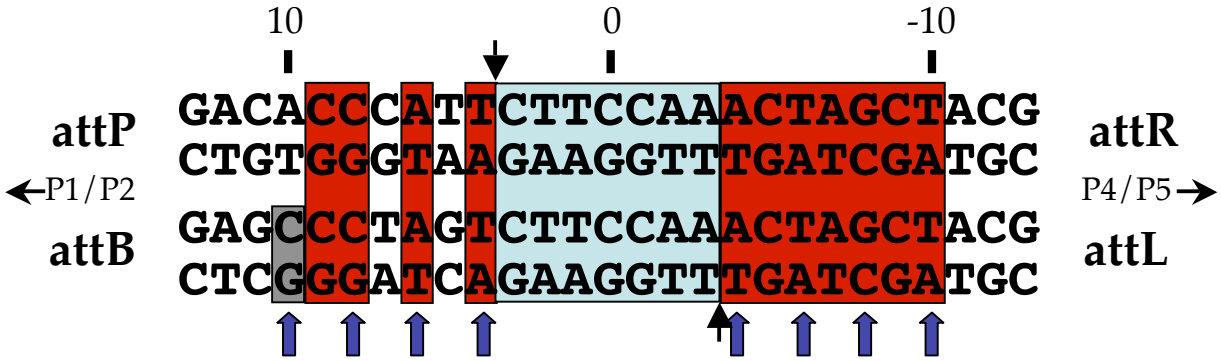
III.H. Discussion

The formation of an active recombinagenic complex requires that L5 Int form a highly regulated complex with *attP*. Int must first bind *attP* in the proper orientation forming the appropriate intermolecular and intramolecular bridges with *attP* and *attB*. Bonds formed between Int and arm-type and core binding sites must then be stabilized so that the complex does not fall apart during recombination.

P1, P2, P3, and P4 only differ in three base pairs collectively (Figure 23). However, Int is able to recognize the differences between these sites and bind P4/P5 selectively during the first step in the synaptic complex building pathway. P4/P5 are clearly bound more tightly by Int than P1/P2 since P1/P2 containing *attR* is unable to complex with Int in the absence of Xis while Int binds P4/P5 containing *attL* readily (Lewis and Hatfull, 2000). After binding P4/P5, Int winds *attP* around a complex of 2 or 4 Int molecules and binds *attP* core at the opposite end of the protein. Bending of

Figure 23. Alignment of RBEs

The sequences of *attP* and *attB* as well as *attR* and *attL* are shown here as indicated with arrows to the right and left showing the location of arm-type binding sites. The 7 base pair overlap region is shown in the light blue box with small black arrows showing the sites of nicking by Int. Base pairs in red boxes show the bases that are shared by left or right RBE. Blue arrows indicate base pairs that make up the imperfect inverted repeats. The final base pair of the inverted repeats are missing in *attB*/*attL* as designated by the gray box.


10
0
-10
attP
GACACCCATTCTTCCAAACTAGCTACG
CTGTGGGTAAGAAGGTTTGATCGATGC
attR
←P1/P2
P4/P5→
attB
GAGCCCTAGTCTTCCAAACTAGCTACG
CTCGGGATCAGAAGGTTTGATCGATGC
attL

attP is stabilized by mIHF, but Int must keep a firm grip on P4/P5 so that arm-type binding is not lost during complex formation. Once core has been delivered to the catalytic domain, Int binds core so that the 7 base pair overlap region is aligned properly for nucleophilic attack. This is accomplished through interactions with the right and left side RBEs that flank the overlap region. At the same time, Int must recognize the difference between *attP*, *attB*, *attL*, and *attR* cores, so that inappropriate complexes are not created.

Arm-type binding sites appear to be used in recognition of *attL* and *attR*. Since *attL* and *attR* contain both core and arm-type binding sites, Int binds *attL* and *attR* bivalently. Int cannot bind *attR* and *attP*, and *attL* and *attP* simultaneously as they will compete for arm-type binding sites in the synaptic complex even if Int protomers are binding both potential substrates at their cores (Figure 15). This indicates that Int cannot distinguish between *attB* and *attP* based on the sequence of the core, but that arm-type binding sites associated with these substrates are the primary determinant for core binding.

The 7 base pair imperfect inverted repeats that make up the RBEs of *attP* core are necessary for core recognition and, therefore, integration. When the RBEs are mutated, recombination does not occur, and the complexes observed when these substrates are used in EMSA are limited to those containing intermolecular bridging between Int and *attB*. The *attP* of pCWCM1 contains mutations in the base pairs of the inverted repeats of the RBEs and not in the intervening base pairs. Surprisingly, This substrate is less proficient at creating stable complexes with Int and *attB* than pCWCM2 that has mutations in all 7 of the RBE base pairs. This could indicate that Int is making interactions with the mutant core that is competing with *attB* for binding causing formation of unsound intasomes that run as smears during EMSA. Therefore, Int must

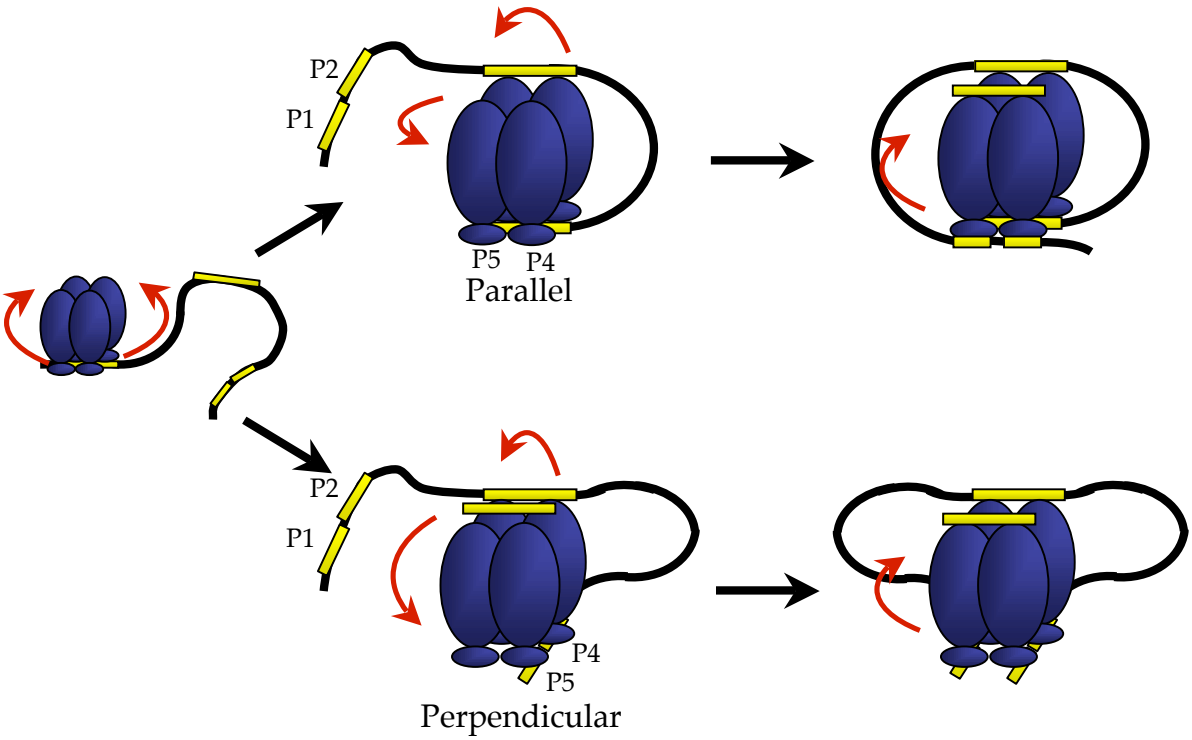
be using these base pairs to recognize the core, but they are not sufficient to create stable Int/core interactions.

During core recognition, Int contacts the right side RBE first and forms a stable contact here. Mutant *attPs* that have been mutated in the right side RBE while the left most RBE remains wild type, pCWCMR.5, cannot be used as substrates for recombination and do not form complexes that require Int/core interactions. However when only the left side RBE is mutated, inefficient recombination occurs, and complexes containing Int/core interactions, intasome and synaptic complex 1 can be created at low frequency. This could indicate that Int binds the right side RBE with greater affinity. The right side RBEs are identical in *attP* and *attB* and, therefore, contain the same intervening base pairs. Since these base pairs appear to interact with Int during core binding, Int could bind the right side RBE with greater affinity based on the sequence of the intervening base pairs.

Binding to P4/P5 activates core binding within Int protomers, however Int can bind core either in parallel, with both Int protomers binding the *attP* core, or perpendicularly, with a single proteomer binding *attP* core while the other proteomer binds *attB* (Figure 24). Parallel binding would allow the activation of the core-binding domain following binding of P4/P5, allowing Int to bind *attP* core. Core binding could cause the activation of neighboring unbound protomers preparing them to bind to P1/P2 and *attB*. Binding *attP* perpendicularly would allow *attB* capture to occur while the synaptic complex is built. The binding of left side RBE could activate the N-terminal domain within the same proteomer, increasing this protomers affinity for P1/P2. P1/P2 binding would activate the remaining proteomer allowing the left side RBE of *attB* to be bound at its catalytic domain. In this model, activation travels between protomers at Ints C-terminus, and within protomers at its N-terminus.

Figure 24. Models of L5 Int Complex Building

Int can form the recominagenic complex through two pathways. If core binds Int in parallel, P4/P5 and core would bind to the same Int protomers. This would leave two unbound Int protomers to bind P1/P2 and *attB*. Red arrows illustrate activation of Int. If *attP* binding activates Int within the same protomer, binding of P4/P5 to Ints N-terminal domain would activate the C-terminal domain for core binding. If Int binds core perpendicularly, the activation of the C-terminal domain would allow binding of both *attP* core and *attB*. The right side RBE binds first activating the neighboring Int protomers C-terminal domain for binding to the lower affinity left side RBE of *attP* core. Binding of the left side RBE activates the N-terminal domain of this protomer, so P1/P2 can bind to this protomer. Binding of P1 to the next protomer activates the final Int protomer's C-terminal domain allowing the lower affinity RBE of *attB* to bind. The recombinagenic complex is formed, and recombination can go forward.


P2
P1
P5
P4
Parallel
P2
P1
P4
P5
Perpendicular

The $tRNA^{gly}$ gene that houses *attB* cannot tolerate mutation. Having the high affinity site on the right of *attP* and *attB* would help protect the insertion site from random mutation since integretion is less likely to occur when mutation occurs here. The presence of a high affinity site on the right side of the 7 base pair overlap region would also explain the inability of *attR* to remain in complex with Int after recombination. If binding to a high affinity site initially is important for the stabilization of the Int/*attP* core interaction, Int may be unable to bind to *attR* core after binding to its arm-type sites since initial contact would be made with a lower affinity RBE. Like pCWCMR.5, the Int core interaction would be too weak for recombination to continue and the complex would fall apart. Equal and increased affinity for one half of the *attP* and *attB* core half sites, allows Int to bind the substrates sequentially during the complex formation pathway increasing the efficiency of the complex building process while minimizing the formation of erroneous complexes

IV. CHARACTERIZATION OF L5 INTEGRASE MUTANTS

IV.A. Introduction

Mycobacteriophage L5 Integrase (Int) is a member of the tyrosine family of recombinases and catalyzes integration by forming intramolecular bridges within the phage attachment site, *attP* (Peña et al. 1997), between the core binding site and a pair of arm-type binding sites to the right of core (P4/P5) (Peña and Hatfull 2000, Peña et al. 1999). Int also mediates intermolecular bridging also between arm-type binding sites P1/P2 of *attP* and the bacterial attachment site, *attB*, on the host chromosome (Peña et al. 1996). A host encoded DNA binding protein, mIHF, binds between the core and arm-type binding sites and is required for recombination (Pedulla and Hatfull 1998, Pedulla et al. 1996).

Int is made up of three domains; the C-terminal core binding region contains the catalytic domain (Hickman et al. 1997, Tirumalai et al. 1997), the N-terminal domain mediates binding to the arm-type binding sites (Segall and Nash 1993), and the middle domain that has been implicated in mediating cross talk between the core and arm-type binding domains, binding directly to core (Patsey and Bruist 1995), and forming protein/protein interactions with mIHF and excisionase (Xis), the phage encoded recombination directionality factor (RDF), which induces excision of the phage genome (Wu et al. 1998).

The catalytic domains of other tyrosine recombinases have been well studied. The catalytic residues have been identified as a tyrosine nucleophile at the extreme C-terminus with an arginine, histidine, arginine triad of residues spread throughout the domain and act to position the substrate in the active site (Gopaul and Duyne 1999). Structural studies have provided a great deal of information regarding how these

residues function together in recombination (Van Duyne 2002). The structure of the N-terminal domain has been solved using NMR, and contains a helix-turn-helix binding motif consistent with the DNA binding domains of other proteins (Wojciak et al. 2002). Besides binding the arm-type sites on *attP*, the N-terminal domain has been implicated in forming protein/protein interactions between protomers of Int (Jessop et al. 2000) and between Int and IHF (Segall et al. 1994), and Int and Xis (Cho et al. 2002, Thompson et al. 1987a) in the lambda recombination system. The binding of xis with λInt increases Int's affinity for the arm-type binding sites, stabilizing excision complexes (Cho et al. 2000, Cho et al. 2002).

The middle domain remains more mysterious. No crystal or NMR structures have been determined for this domain of any tyrosine recombinase, and sequence comparisons of various tyrosine recombinases within this region show very little similarity. The position of this domain between the catalytic and arm-type binding domains would indicate that signals between the outer domains are carried through this domain (Han et al. 1994b). It has been suggested that this domain is utilized in binding to *attP* core (Dorgai et al. 1998, Han et al. 1993, Tirumalai et al. 1998). Still other reports suggest that this domain forms protein/protein interactions with host factors required for recombination (Cho et al. 2002, Segall et al. 1994, Thompson et al. 1987a). In all, very little relevant information exists concerning the function of this domain, but it's location and the complexity of the integrative complex suggest that its proper function is essential for integrative recombination.

IV.B. Specific Aims

1. Isolate L5 Int mutants that contain single base pair changes and therefore single amino acid substitutions in the middle domain

2. Biochemically characterize these mutants while comparing them to wild type L5 Int to help discern the function of the middle domain.

IV.C. Creation of Middle Domain L5 Integrase Mutants

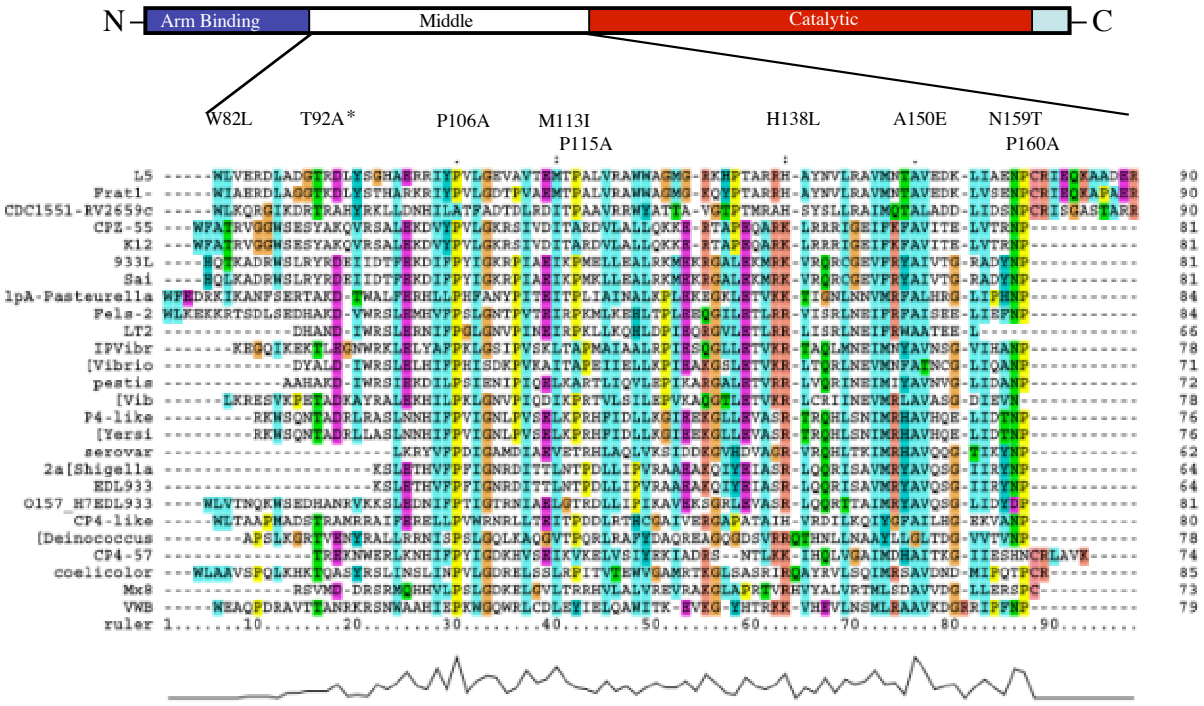
Characterization of the middle domain of L5 Int was carried out using sequence comparisons. A list of middle domains of tyrosine recombinases closely related to L5 Int was compiled using BLAST (Altschul et al. 1990) and PSI-BLAST (Altschul et al. 1997) algorithms with each recombinase being used as a query sequence. Non-tyrosine recombinase matches were unusual, and these proteins were eliminated. The recombinase middle domains were then aligned using Clustal X (available from BioWeb at http://www.web.tiscalnet.it/biologia). Conserved amino acids were those that appeared in several aligned Ints at the same location in the sequence (Figure 25). Site directed mutants were created at conserved residues, 82, 92, 115, 138, 150, 159, and 160 using the Stratagene QuickChange mutagenesis kit on pPSC1, a pET21a derived expression vector containing L5 Int. The entire *Int* gene for each mutant was then sequenced to ensure that the proper base pair had been replaced and that no other mutations had occurred anywhere within the *Int* gene during the mutagenesis process. Mutants created are described in Table 2.

IV.D. Initial Characterization of Middle Domain L5 Integrase Mutants

Each middle domain mutant plasmid described above was transformed into BL21(DE3)pLysS *E. coli* cells. Bacteria containing the mutant plasmids were selected by resistance to carbenicillin. Individual colonies were picked and grown in liquid culture to A_{600} of 0.6-0.8 at 30° C. These cultures were then induced with 0.1 mM Isopropyl □-$_D$-thiogalactopyranoside (IPTG) and allowed to continue growth for 2.5 hours. Lysates

Figure 25. Alignment of Middle Domains of Closely Related Tyrosine Recombinases

A linear representation of L5 Int is shown above the alignment. The catalytic domain is colored red, the arm-type binding domain blue, and the middle domain white. The list of tyrosine recombinases closely related to L5 Int was generated using BLAST and 2 iterations of Psi-BLAST, and Clustal X was used to align these recombinases. Color is used in the alignment to group amino acids based on charge and size. Conserved amino acids are highlighted by marks above the residue. Below the alignment is a line graph representing the similarity of the amino acids. Spikes in the line represent more conserved amino acids. Conserved residues W82, T92, P106, M113, P115, H138, A150, N159, and P160 were chosen for mutagenesis.


N
Arm Binding
Middle
Catalytic
C
W82L
T92A*
P106A
M113I
P115A
H138L
A150E
N159T
P160A
L5
Fratl-
CDC1551-RV2659c
CPE-55
K12
933L
Sai
lpA-Pasteurella
Pels-2
LT2
IPVibr
[Vibrio
pestis
[Vib
P4-like
[Yersi
serovar
2a[Shigella
EDL933
O157_H7EDL933
CP4-like
[Deinococcus
CP4-57
coelicolor
Mx8
VWB
ruler 1.....10.....20.....30.....40.....50.....60.....70.....80.....90.....

containing mutant Ints were harvested by freezing pelleted cultures at –80° C and thawing on ice with room temperature buffer. SDS-PAGE was used to confirm that mutant proteins were expressed to satisfactory levels (Figure 26).

L5 Int forms several distinct complexes both on ice and at room temperature that can be distinguished using Electrophoretic Mobility Shift Analysis (EMSA) (Pena et al. 2000, Peña and Hatfull 2000, Peña et al. 1999). Since the ability to form these complexes is a defining feature of L5 Int, EMSA was utilized in the characterization L5 Int middle domain mutants. One μl of each mutant Int extract was mixed with radio labeled *attP*, 289 nM *attB*, and 140 nM mIHF. Complexes were allowed to form on ice for 20 minutes, or complexes were formed on ice for 20 minutes before being shifted room temperature for 20 minutes. The resulting complexes were separated by electrophoresis on a 5% non-denaturing poly-acrylamide gel at 4°C (Figure 27). At least four different phenotypes were observed in gel shift analysis. P115A, P160A, and N159T mutants form complexes, as well as small amounts of product similar to wild type Int. W82L, T92A, and H138L formed consistent with the intasome, a complex containing Int and mIHF in which intramolecular bridges are formed with Int between core and P4/P5 of *attP*. None of these mutants formed synaptic complex 1 wherein Int bridges P4/P5 and core with the help of mIHF and P1/P2 and *attB*, or synaptic complex 2 in which Int forms intermolecular bridges between P1/P2 and *attB*. W82L was also found to create a small amount of product (less then wild type) when incubated at room temperature. A150E did not form distinct complexes, but could bind *attP* as indicated by smears near the substrate band. Because their mutant phenotypes were more severe, W82L, T92A, H138L, and A150E were selected for further analysis.

Figure 26. Overexpression of Middle Domain Int Mutants

Site-directed middle domain mutants and wild type L5 Int were overexpressed using a pET21a derivative in BL21(DE3)pLysS cells. Coomasie Brillant Blue stained SDS-PAGE was used to show the expression of each of the proteins. Lane 1 contains standard protein makers (Bio-Rad). Purified wild type Int is shown in lane 2 followed by an uninduced sample. Lanes 5-8 contain total protein lysates of wild type Int, T92A, P106A, A150E, and P160A.

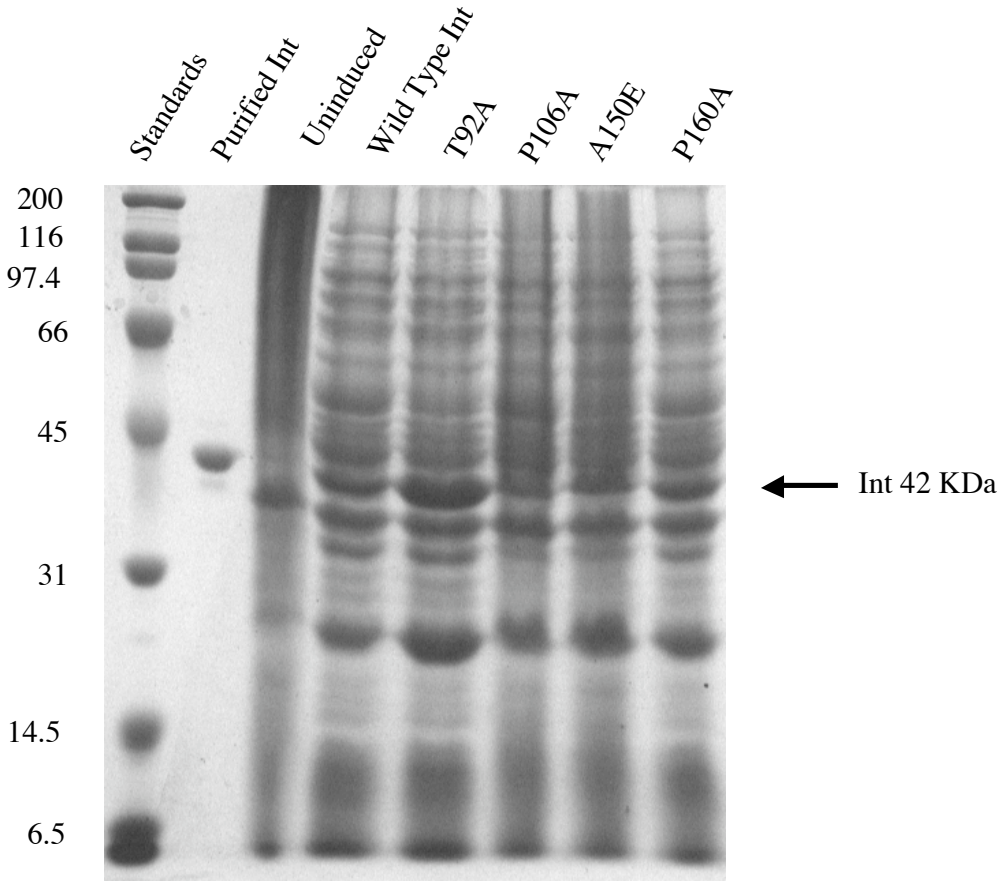
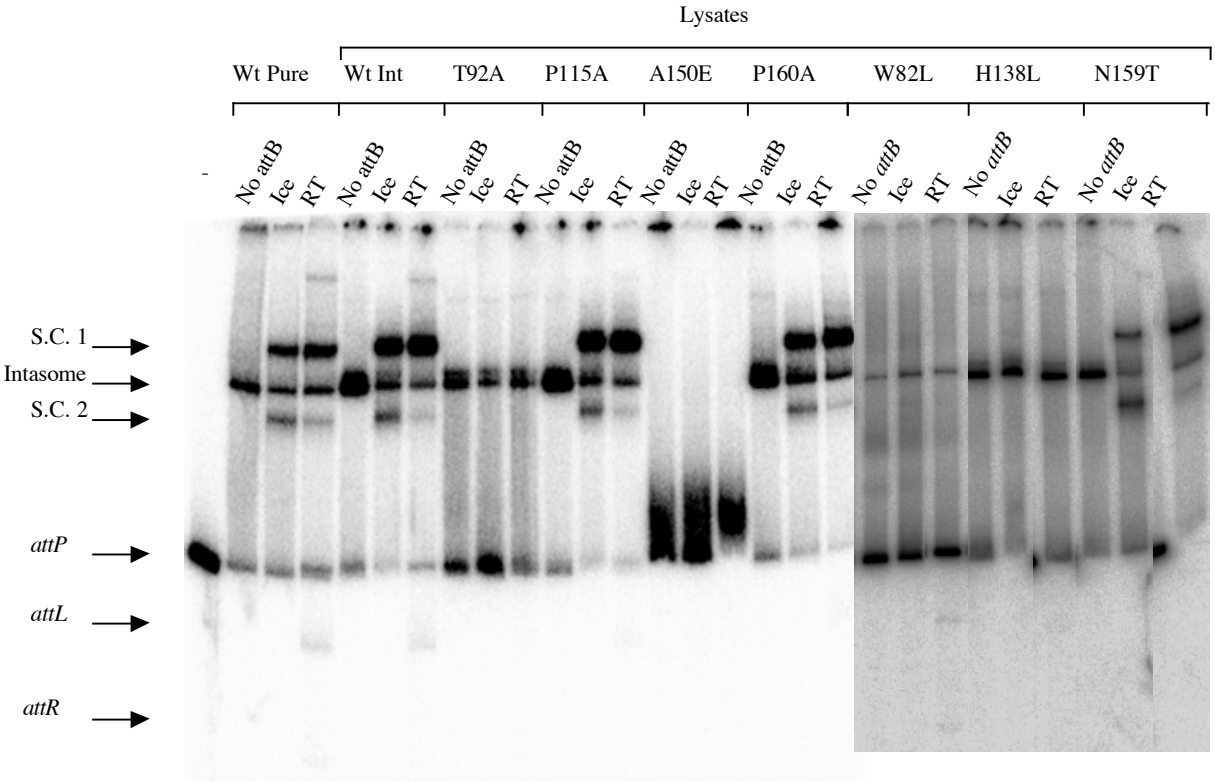

Standards
Purified Int
Uninduced
Wild Type Int
T92A
P106A
A150E
P160A
200
116
97.4
66
45
31
14.5
6.5
Int 42 KDa

Figure 27. EMSA of Middle Domain Int Mutants

One μl of each mutant lystate was mixed with 560 nM mIHF, 140 nM *attB*, and radio labeled *attP* under standard reaction conditions. Complexes were allowed to form for 20 minutes on ice or were formed on ice for 20 minutes and shifted to room temperature for 20 minutes. Complexes were separated by electrophoresis on a 5% non-denaturing poly-acrylamide gel at 4ºC and visualized by autoradiography. Wild type Int forms intasome and synaptic complexe 1 that runs above intasome and synaptic complex 2 observed below intsaome when complexes are formed on ice. These complexes are also formed at room temperature although the intensity of these bands is different. The recombination products, *attL* and *attR* run below the substrate band and are observed only at room temperature. Mutants P115A, N159T and P160A form complexes like wild type Int and product at room temperature. W82L, T92A and H138L form Intasome but neither synaptic complex. A faint *attL* band is also noted when W82L is incubated room temperature. A150E is unable to form any complexes under these conditions, but appears to be able to bind to *attP* based on smears near the substrate band.


Lysates
Wt Pure
Wt Int
T92A
P115A
A150E
P160A
W82L
H138L
N159T
No attB
Ice
RT
No attB
Ice
RT
No attB
Ice
RT
No attB
Ice
RT
No attB
Ice
RT
No attB
Ice
RT
No attB
Ice
RT
No attB
Ice
RT
S.C. 1
Intasome
S.C. 2
attP
attL
attR

IV.E. Purification of L5 Int Middle Domain Mutants W82L, T92A, H138L, and A150E

L5 Int mutants W82L, T92A, H138L, and A150E were purified using the L5 Int purification methods previously described (Lee and Hatfull 1993). Each mutant was expressed as described above. Ammonium sulfate precipitation was used to eliminate some of the E. coli proteins, and lysates enriched for Int were separated on a carboxymethyl (CM) column using a BioCad Sprint System and fractions containing the muatant Ints were collected. A POROS PE20 column was then used to further purifiy the Int containing fractions leaving a solution containing high levels of the mutant Ints with only minor contaminants. The purity of the mutant proteins was determined using SDS-PAGE, and the concentration of the proteins was quantified using Bradford assay. Each mutant protein was then subjected to band shift analysis as described above to verify its mutant phenotype. *In vitro* recombination assays were then used to assess the mutant Int's ability to perform recombination on supercoiled and linear substrate (Figure 28 A and B). *In vitro* recombination assays were carried out by mixing 1 μl of 1/10, 1/30 and 1/100 dilutions of each purified mutant protein with 70.0 nM *attB* and 280 nM mIHF and incubating for 2 hours at 37ºC. The reactions were heat killed at 65ºC for 20 minutes, and he products were then separated by electrophoresis on a 1% agarose gel and visualized using ethidium bromide staining under UV light.

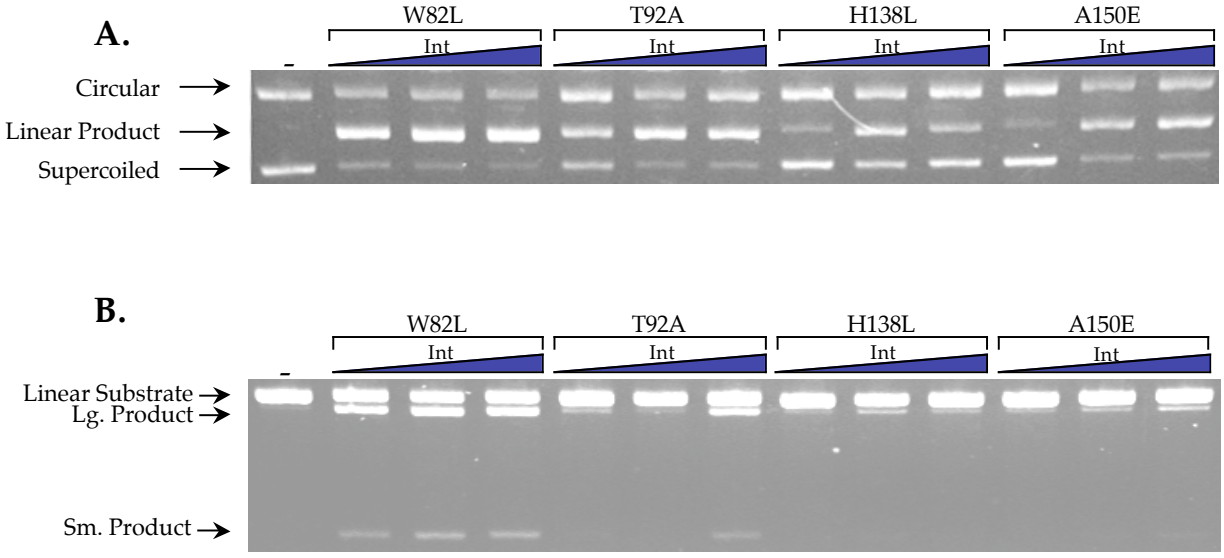
Despite their inability to form the synaptic complexes observed by EMSA in wild type Int, all of the middle domain mutants retain the ability to perform recombination *in vitro*. The ability to perform recombination in the absence of what were believed to be functional intermediates in the recombination reaction indicates that these mutants either perform recombination using a different pathway than wild type L5 Int, the observed complexes are artifacts of EMSA and not really intermediates, or that the

Figure 28. ***In vitro*** **Recombination using Purified Middle Domain Int Mutants**

One μl of 1/10, 1/30, and 1/100 dilutions of each purified mutant Int was mixed with 280 nM mIHF, 70 nM *attB* and 3.40 nM supercoiled *attP* under standard *In vitro* reaction conditions. Recombination was allowed to take place for 2 hours at 37ºC. The reactions were stopped by heating to 65ºC for 20 minutes, and products were separated on a 10% agarose gel and visualized by ethidium bromide staining under UV light.

A. *In vitro recombination on supercoiled substrate*. When supercoiled substrate is used, linear plasmid is produced as a result of recombination. Supercoiled substrate runs below the linear product, and a faint band of nicked circular product runs above the linear product. Each of the middle domain mutants is competent to perform recombination on supercoiled substrate.

B. *In vitro recombination on linear substrate*. Linear substrate was also utilized to test for recombination. Digesting pCPDR13 with XmnI created linear substrate. The result of recombination on this substrate is a 5899 base pair band that runs just below the substrate band and a smaller 724 base pair band. All of the middle domain mutant Ints were able to perform recombination on linear substrate.


A.
W82L
T92A
H138L
A150E
Int
Int
Int
Int
–
Circular
Linear Product
Supercoiled
B.
W82L
T92A
H138L
A150E
Int
Int
Int
Int
–
Linear Substrate
Lg. Product
Sm. Product

synaptic complexes are significantly less stable then the wild type L5 Int complexes and cannot be separated using native gel electrophoresis.

IV.F. Kinetic Analysis of L5 Int Mutants

Since L5 Int mutants, W82L, T92A, H138L, and A150E, are able to perform recombination *in vitro*, kinetic analysis of these mutants using supercoiled or linear *attP* with a linear 42 base pair *attB* as substrate, as well as using supercoiled *attB* with linear 342 base pair *attP* was performed to determine the relative rates of recombination in comparison to wild type L5 Int. *In vitro* reactions were carried out using equimolar amounts of each mutant protein (11.9 nM) in separate reactions containing each of the substrates, *attP*, *attB* as described above, and 280 nM mIHF. Recombination was allowed to continue for various amounts of time (5 – 120 minutes) before being stopped by heating at 65°C for 20 minutes. L5 Int remains bound to *attL* the leftmost product of integration after recombination is completed, therefore the resulting rates of reaction represent single turnover reactions. Products were then separated by agarose gel electrophoresis and visualized under UV light using ethidium bromide staining (Figure 29 A, B, and C). The density of the product band was quantified using a Bio-Rad Gel Documentation system and QuantityOne software. The density of the product band was then plotted versus the reaction time, and a line was fitted to the results (Figure 5 D). Recombination rates are assembled in Table 3.

Mutant Ints perform recombination with supercoiled *attP* at slower rates than wild type Int. The maximum rate of recombination for wild type Int using supercoiled *attP* and linear *attB* as substrates was determined to be 23.9 per minute. The rates of recombination for mutants W82L and T92A are less than 60% of the wild type rate at

Figure 29. Representative Data for Kinetic Analysis of Middle Domain Int Mutants

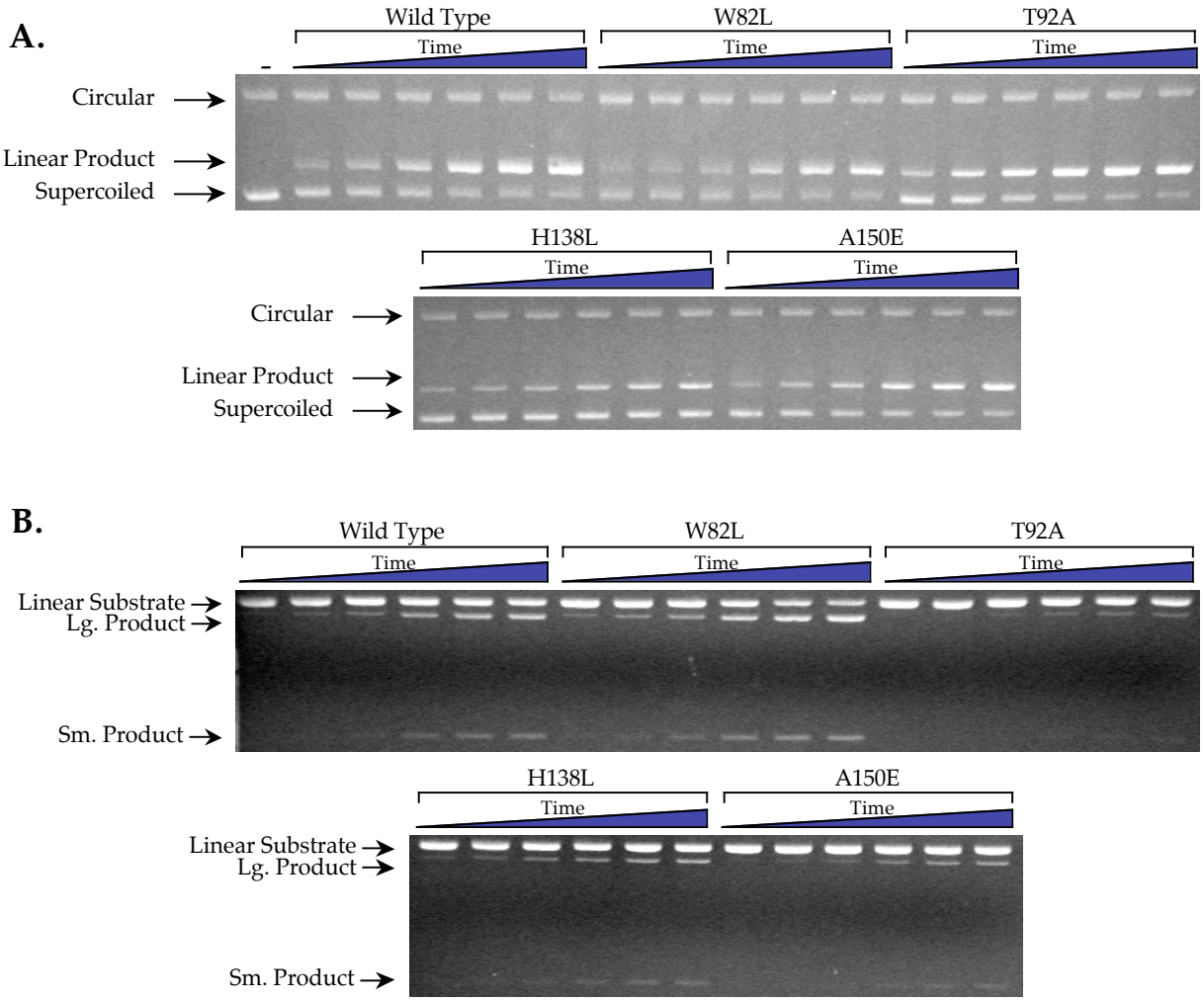
In vitro recombination was used to determine the rates of reactions for each of the middle domain mutants. Reactions containing 11.9 nM Int, 560 nM mIHF, 140 nM linear substrate and 3.4 nM supercoiled substrate were carried out under standard reaction conditions at 37°C for 5, 15, 30, 60, 90, or 120 minutes. The products were run on a 1% agarose gel, stained with ethidium bromide and visualized under UV light. The resulting product bands were quantified using BioRad Gel Documentation system and QuantityOne software. The background of each lane was subtracted from density of each band, and these data were plotted versus time. The resulting plots were fit to a line using pro Fit 5.6.0 (avialible from http://www.quansoft .com/) using 10,000 iterations of Monte Carlo algorithm.

A. *Representative data using supercoiled attP.* 5899 base pair, *attP* containing pCP□R13 and 42 base pair linear *attB* were used as substrates for *in vitro* recombination. The linear product of recombination runs between a faster moving supercoiled substrate band and a slower moving nicked circular substrate band. As the reaction time is increased, the intensity of the product band increases, and since supercoiled substrate is preferred, the intensity of the lower substrate band decreases as more linear product is produced.

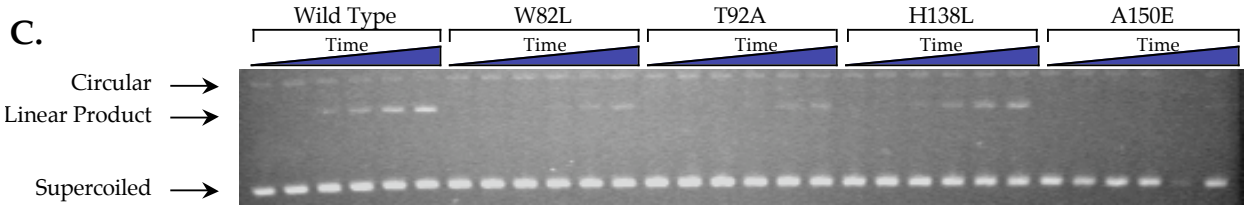
B. *Representative data using linear substrate*. A 5899 base pair linear *attP* created from pCP□R13 that has been cut with XmnI was used as substrate with a 42 base pair linear *attB* in *in vitro* recombination assays. The products of recombination are a 5155 base pair band that runs just below the substrate band that increases in intensity as a function of time and a smaller 734 base pair band.

C. *Representative data using supercoiled attB and linear attP*. A 3186 base pair, *attB* containing plasmid (pJL38) and a 342 base pair BamHI, EcoRI fragment of pCP□R13 containing *attP* were used as substrates for *in vitro* recombination. The linear product of recombination increases in intensity as a function of time and runs between supercoiled pJL38 and nicked circular pJL38.

D. *Representative plots of product intensity versus time*. Data generated through the determination of the intensity of the product bands were plotted versus time and fit to a line using ProFit 5.6.0. The slope of these lines represents the rate of a single turnover recombination reaction per minute for each mutant Int.


A.
Wild Type
W82L
T92A
Time
Time
Time
Circular
Linear Product
Supercoiled
H138L
A150E
Time
Time
Circular
Linear Product
Supercoiled
B.
Wild Type
W82L
T92A
Time
Time
Time
Linear Substrate
Lg. Product
Sm. Product
H138L
A150E
Time
Time
Linear Substrate
Lg. Product
Sm. Product

C.


Wild Type
W82L
T92A
H138L
A150E
Time
Circular
Linear Product
Supercoiled

D.

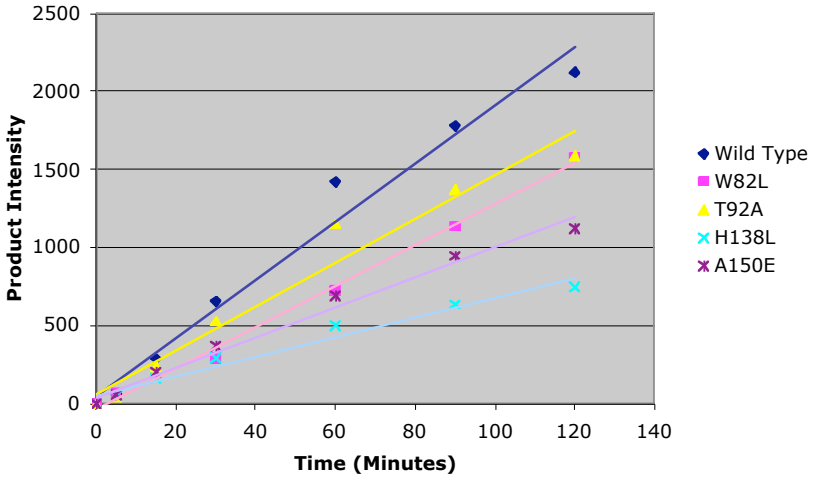

Product Intensity
2500
2000
1500
1000
500
0
0
20
40
60
80
100
120
140
Time (Minutes)
Wild Type
W82L
T92A
H138L
A150E

Table 3. Rates of Reaction For Middle Domain Int Mutants

Substrate	Wild Type	W82L	T92A	H138L	A150E
Supercoiled *attP*	23.9±2.74 (100%)*	13.5±2.50 (56.6%)	14.1±3.83 (59.1%)	8.33±3.21 (34.9%)	9.70±0.42 (40.6%)
Linear *attP*	16.7±4.34 (69.9%)	16.6±3.79 (69.6%)	2.51±1.28 (10.5%)	6.51±2.82 (27.2%)	3.38±1.64 (14.1%)
Supercoiled *attB*	4.40±0.14 (18.4%)	1.35±0.21 (5.65%)	1.30 (5.44%)	4.50±2.26 (18.8%)	0.20±0.14 (0.84%)

*Rates ± Standard Deviation (Percent Rate Supercoiled *attP*)

13.5 per minute and 14.3 per minute, respectively. H138L and A150E perform recombination at even lower rates, 8.33 per minute or 34.9% wild type and 9.70 per minute or 40.6% wild type, respectively.

Wild type Int can recombine linear *attP* and *attB in vitro* at 16.7 per minute or 69.9% wild type supercoiled. T92A, H138L and A150E, performed recombination much less efficiently on linear substrates at 2.51 per minute or 10.5% of the wild type linear rate for T92A, 6.51 per minute or 27.3% of the wild type linear rate for H138L, and 3.38 per minute or 14.13% of the wild type linear rate for A150 E. Surprisingly, W82L performed recombination as efficiently as wild type Int at 16.6 per minute.

In vitro recombination of linear *attP* with supercoiled *attB* substrates occurs much less efficiently than recombination of supercoiled *attP* and linear *attB* substrates. Wild type Int performs recombination at 18.4% the rate recombination of supercoiled *attP* and linear *attB* by wild type Int or at 4.40 per minute. W82L, T92A, and A150E perform recombination at less efficiently with supercoiled *attB* and linear *attP* at 1.35 per minute or 30.7% the rate wild type Int, 1.30 per minute or 29.6% the rate of wild type Int, and 0.2 per minute or 4.55% the rate of wild type Int, respectively, using supercoiled *attB* as substrate. H138L performs recombination at the same rate as wild type Int at 4.50 per minute using supercoiled *attB* and linear *attP* as substrates.

Recombination rates vary quite drastically for wild type Int depending on the substrate. Rates decrease 30% when supercoiled *attP* is exchanged for linear *attP* as substrate with linear *attB*. The decrease is even more dramatic when supercoiled *attB* and linear *attP* are used as substrates, almost 80%. Mutants T92A and A150E utilize these substrates with similar rate deductions. W82L recombines supercoiled *attP* and linear *attP* with linear *attB* with similar efficiency and performs recombination as

efficiently as wild type Int on linear *attP* and *attB*. There is a greater then 50% decrease in reaction rate when supercoiled *attB* and linear *attP* are used as substrates. This could be due to the relative small size of the *attP* substrate, 342 base pairs. L5 Int has an intrinsically high affinity for non-specific DNA. W82L may bind to the non-specific plasmid DNA that makes the supercoiled *attB* substrate, and may have trouble capturing a small *attP* molecule. Interestingly, H138L shows very little preference for substrate and catalyzes recombination with with all of the substrates tested at similar rates.

IV.G. The Effect of mIHF Concentration on Mutant L5 Int *In vitro* Recombination

To test the effect of mIHF concentration on recombination, *in vitro* recombination reactions were carried out using 3.4 nM linear *attP* and 70 nM linear *attB* as substrate in the presence of increasing concentrations of mIHF (7 nM – 2.8 μM). Reactions were incubated for 2 hours at 37ºC. Int begins to lose activity after 2 hours at 37ºC since no product formation is observed even when substrate is presents after 2 hours. The products were separated using agarose gel electrophoresis and visualized by ethidium bromide staining under UV light (Figure 30 A). The results were then quantified using Bio-Rad Gel Documentation system with QuantityOne software. The product density was determined and divided by the total density of the DNA, which was multiplied by 100% to give the percent of total DNA that has been converted to product. The resulting percent total DNA was then plotted versus mIHF concentration (Figure 6B).

The increased intensity of the large product band at high mIHF concentrations shows that the mutant Ints are able to perform recombination more efficiently at high mIHF concentration. Wild type Int performs recombination best at 560 nM mIHF under

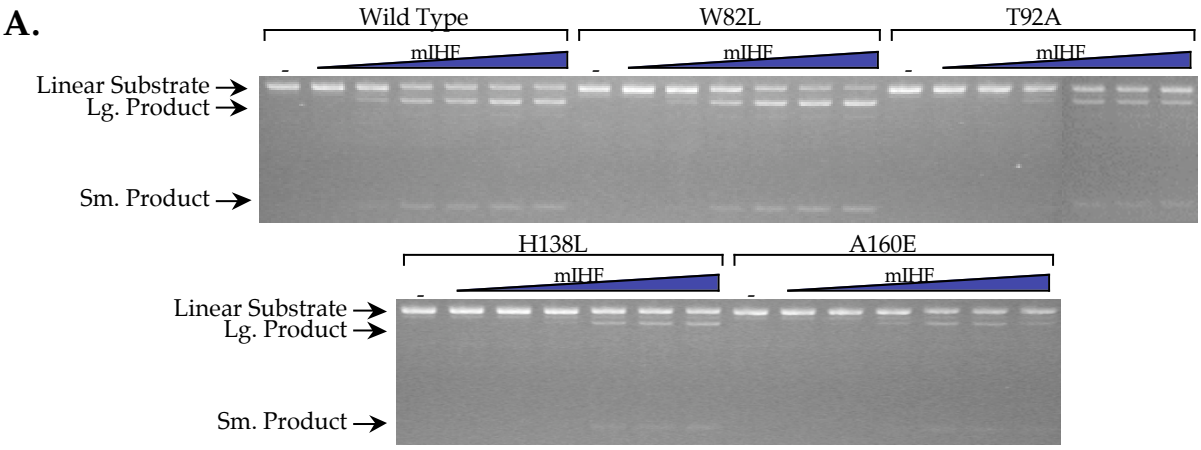
Figure 30. The Effect of mIHF Concentration of *in vitro* Recombination: Linear *attP*

In vitro recombination was carried out on 3.4 nM linear 5899 base pair *attB* using 11.9 nM Int, 70nM 42 base pair *attB* and increasing concentrations of mIHF (7 nM – 2.8 μM) under standard reaction conditions. Recombination was allowed to occur at 37°C for 2 hours before being stopped by heating to 65°C for 20 minutes. Products were separated using agarose gel electrophoresis on 1% agarose gel and visualized by ethidium bromide staining under UV light. The results were quantified using BioRad Gel Documentation system with QuantityOne software.

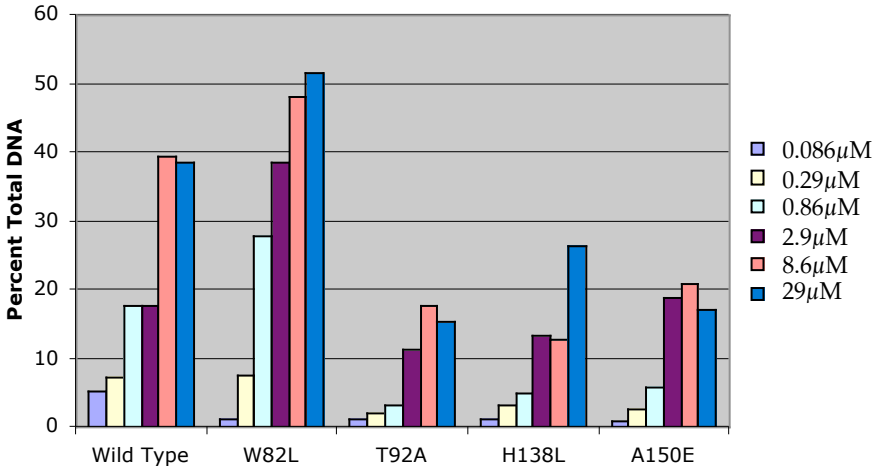
A. *Representative data using linear substrates with increasing mIHF concentrations.* A 5155 base pair band that runs just below the substrate band and 734 base pair band are the result of recombination of linear substrates. The concentration of mIHF in the reaction is increased from left to right. The increased intensity of product bands at high mIHF concentrations shows that recombination occurs more efficiently at high concentrations for middle domain Int mutants whereas high mIHF concentrations appears to inhibit recombination in reactions containing wild type Int.

B. *Plot of the percent of the DNA converted to product versus the mIHF concentration*. The density of the larger 5155 base pair product was determined using QuantityOne software and background was subtracted for each lane. The resulting density was divided by the total density of the DNA in each lane and multiplied by 100 to get the percent total DNA in the product band. This was done for each mIHF concentration several times, and the average was plotted versus the mIHF concentration. The vertical bars represent the standard deviation of the data points compiled.

A.


Wild Type
W82L
T92A
mIHF
-
Linear Substrate
Lg. Product
Sm. Product
H138L
A160E

B.


Percent Total DNA
60
50
40
30
20
10
0
Wild Type
W82L
T92A
H138L
A150E
0.086μM
0.29μM
0.86μM
2.9μM
8.6μM
29μM

standard conditions, but at high concentrations of mIHF (>2.8 μM), recombination is inhibited. W82L performs recombination more efficiently than wild type Int on linear substrates, which is consistent with the kinetic data, and appears to create more product as the concentration of mIHF is increased. W82L performs recombination well even at mIHF concentrations that were inhibitory for wild type Int. Similarly, H138L performs recombination more efficiently at high mIHF concentration. On average, almost twice as much product is observed at 2.8 μM mIHF then at 1.9 μM. A150E and T92A appear to lose efficiency at high mIHF concentrations. Similar results were obtained when supercoiled *attP* was used as substrate (Figure 31 A and B).

L5 Int recombination requires mIHF (Pedulla and Hatfull 1998, Pedulla et al. 1996). These data show that recombination is inhibited by high mIHF concentrations in reactions containing wild type Int. This does not appear to be the case for middle domain mutants, W82L and H138L. These mutants are able to perform recombination more efficiently at high mIHF concentrations indicating that mIHF is able to compensate, at least in part, for these mutants deficiencies in either the binding to substrate DNA or catalyzing of recombination.

IV.H. The Effect of *attP* Spacing Mutations on Mutant L5 Int In vitro Recombination

The number of base pairs between the core binding region, and the arm-type binding sites has been shown to affect both complex formation and the ability of L5 Int to perform recombination *in vitro* (Peña and Hatfull 2000). This is thought to be due to both the face of the DNA helix on which the core binding region resides in relation to the arm-type binding region and the amount of bending required to bring the core binding region into the proper position on the core binding domain of Int. If the complex formation and *in vitro* effects noted for these middle domain mutants

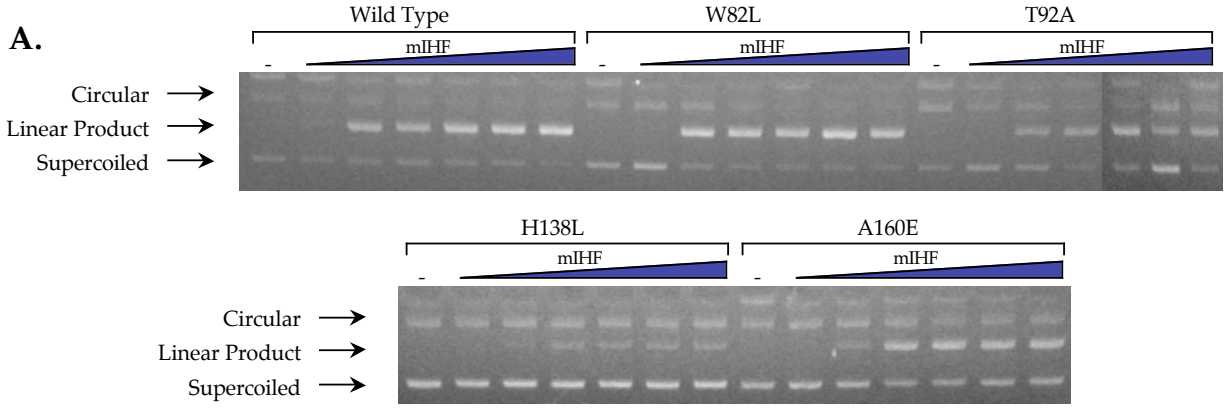
Figure 31. The Effect of mIHF Concentration on *in vitro* Recombination: Supercoiled *attP*

In vitro recombination was carried out under standard conditions using 11.9 nM Int, 70nM 42 base pair *attB*, increasing concentrations of mIHF (7nM – 2.8μM) and 3.4 nM 5899 base pair supercoiled *attP*. Recombination continued for 2 hours at 37ºC before being stopped by heating to 65ºC for 20 minutes. Products were visualized by ethidium bromide staining under UV light. Data was collected using BioRad Gel Documentation system with QuantityOne software.

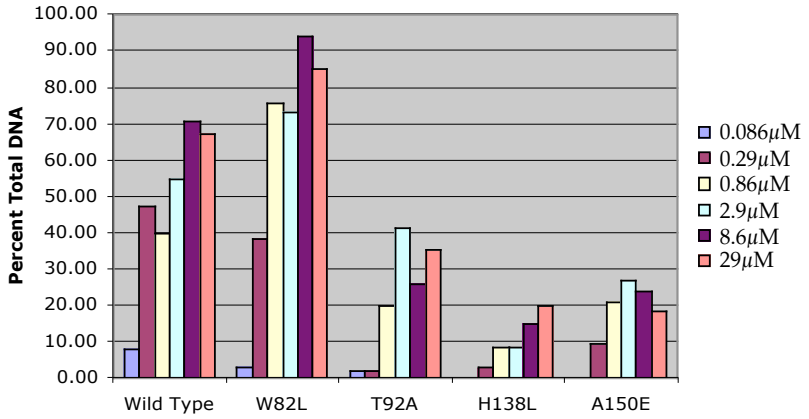
A. *Representative data using supercoiled attP and linear attB and increasing the mIHF concentration*. Linear product runs between a faster moving supercoiled band and a slower running nicked circular product. The concentration of mIHF is increased in each set reactions from left to right. These results show that wild type Int does not perform recombination as efficiently at high concentrations of mIHF. However, middle domain mutants perform recombination better at high concentrations of mIHF.

B. *Plot of the percent total DNA converted to product versus mIHF concentration*. The density of the product band was determined using QuantityOne software, and the background for each lane was subtracted. The resulting densities were then divided by the total DNA substrate and multiplied by 100 to give the percent total DNA converted to product. The average percent after several trials was plotted versus the mIHF concentration in each reaction, and vertical bars represent the standard deviation of these data.

A.


Wild Type
W82L
T92A
mIHF
-
Circular
Linear Product
Supercoiled
H138L
A160E
mIHF
-
Circular
Linear Product
Supercoiled

B.


Percent Total DNA
100.00
90.00
80.00
70.00
60.00
50.00
40.00
30.00
20.00
10.00
0.00
Wild Type
W82L
T92A
H138L
A150E
0.086μM
0.29μM
0.86μM
2.9μM
8.6μM
29μM

represents their reduced ability to bring mIHF into the recombinagenic complex, these effects should be exaggerated in *attP* mutants that have extended DNA sequence between the arm and core binding regions since mIHF is less likely to join these complexes.

To examine the effect of *attP* spacing mutants on *In vitro* recombination, plasmids containing *attP* mutants pMK6 and pMK17 with insertions between the core and P4/P5 arm-type binding sites of 11 and 13 base pair, respectively, were used as substrates with a 42 base pair *attB* in *in vitro* recombination assays with the middle domain mutants. pMK6 has been shown to support recombination *in vitro*, while pMK17 is unable to support recombination (Peña and Hatfull 2000). *In vitro* recombination was carried out as described above, for discrete amounts of time (5 - 120 minutes) before being heat killed at 65ºC for 20 minutes. Products were separated by agarose gel electrophoresis and visualized using ethidium bromide staining under UV light. The amount of product produced was determined using BioRad Gel Documentation system with QuantityOne software. The product produced in each reaction was plotted versus the time of the reaction, and a line was fit to the resulting plot. The relative rates of reactions could then be determined, and compared to wild type.

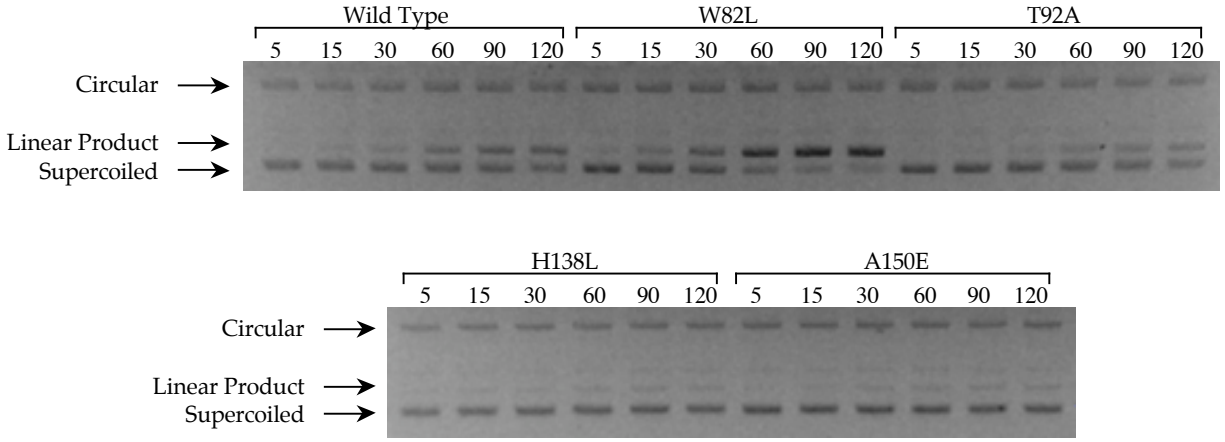
Wild type Int was able to perform recombination on supercoiled pMK6 at approximately 10% the rate of supercoiled wild type *attP*, pCP□R13, or 2.3 per minute (Figure 32). As expected, mutants T92A, H138L and A150E performed recombination on pMK6 much slower than wild type Int on this substrate at 87.4%, 14.7%, and 11.8%. It should be noted that very little product is observed even after 2 hours at 37ºC. Surprisingly, W82L performs recombination much more efficiently than wild type Int on pMK6, 5.5 per minute or 233% the rate of wild type Int.

Figure 32. pMK6 *in vitro* Recombination

A. *In vitro recombination with pMK6*. *In vitro* recombination was carried out using 3.4 nM 7774 base pair pMK6 as the supercoiled *attP* substrate. pMK6 contains an 11 base pair insertion between P4/P5 and core. Reactions were carried out under standard conditions with 11.9 nM Int, 70 nM *attB* and 280 nM mIHF for various times (5 – 120 minutes). Products were separated by agarose gel electrophoresis on a 1% agarose gel. Products were visualized by ethidium bromide staining under UV light and quantified using BioRad Gel Documentation system with Quantity One software.

B. *Rates of Recombination with pMK6.* Rates of reaction were calculated as previously described, and resulting rates are reported here. Rates are given as substrate converted per minute, and percentages are compared to supercoiled wild type *attP* with linear *attB* substrates.

A.



B.

Wild Type	W82L	T92A	H138L	A150E
2.38 (10%)	5.47 (23.2%)	2.07 (8.7%)	0.35 (1.5%)	0.28 (1.2%)

As previously reported, initial time course data suggested that wild type Int was unable to perform recombination on pMK17 even after 2 hours at 37ºC (Peña and Hatfull 2000). Mutants T92A, H138L and A150E were also unable to perform recombination on this substrate at detectable levels. Unlike wild type, W82L could catalyze recombination with pMK17. Very low levels of product can be seen after 90 and 120 minutes of incubation at 37ºC (Figure 33 A). However, the amount of product produced is sufficiently low to prevent the determination of rate constants.

To further examine this phenomenon, *in vitro* recombination was performed using increasing amounts of mIHF. If mIHF binding to *attP* DNA facilitates the DNA bending required to bring *attP* core into close proximity of the Int catalytic domain and substrates with insertions between core and arm-type sites require greater bending of the *attP* DNA to deliver core to the catalytic domain, than mIHF might be limiting in reactions using pMK17 as substrate. Increasing the mIHF concentration might therefore rescue this reaction and allowing the formation of more product.

In vitro reactions were carried out as described above using supercoiled pMK17 and 42 base linear *attB* as substrates, and the mIHF concentration was increased from 0 – 2.8 μM as indicated (Figure 33 B). Wild type Int is able to perform recombination on pMK17 at low levels and more product is observed as the amount of mIHF is increased. This is also true for W82L. As the mIHF concentration is increased, W82L is more capable of performing recombination. Mutants T92A, H138L and A150E were unable to perform recombination at detectible levels at any mIHF concentration tested.

Several other *attP* insertion mutants have been characterized and were used as substrates in *in vitro* reactions to determine if increasing the mIHF concentration in reactions using *attP* insertion mutants that were previously reported to be unable to support recombination can rescue this phenotype. Wild type and W82L were tested for

Figure 33. pMK17 *in vitro* Recombination

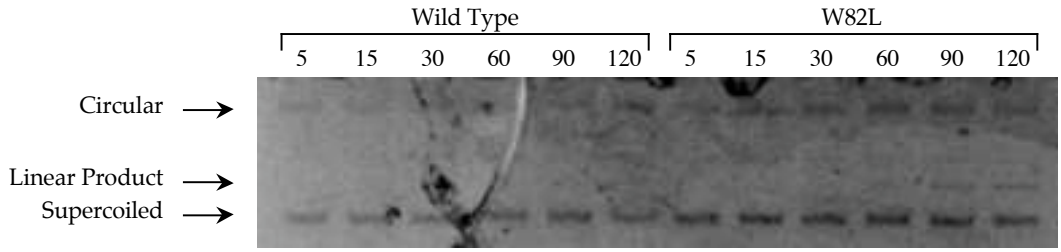
In vitro recombination were performed using 3.4 nM 7776 base pair pMK17 that has a 13 base pair insert between P4/P5 and core, as supercoiled DNA substrate. Reactions were carried out under standard conditions in the presence of 11.9 nM Int, 70 nM 42 base pair *attB*, and 280 nM mIHF. Recombination was performed at 37ºC for 2 hours, and heating to 65ºC for 20 minutes terminated reactions. Products were separated by electrophoresis on a 1% agarose gel and visualized by ethidium bromide staining under UV light. The images have been inverted to aid in the visualization of minute amounts of product.

A. *Time course of in vitro recombination of wild type Int and W82L with pMK17*. Initial time course data shows a small amount of product formed by W82L at 90 and 120 minutes. There does not appear to be any product formed by wild type Int even after 2 hours.

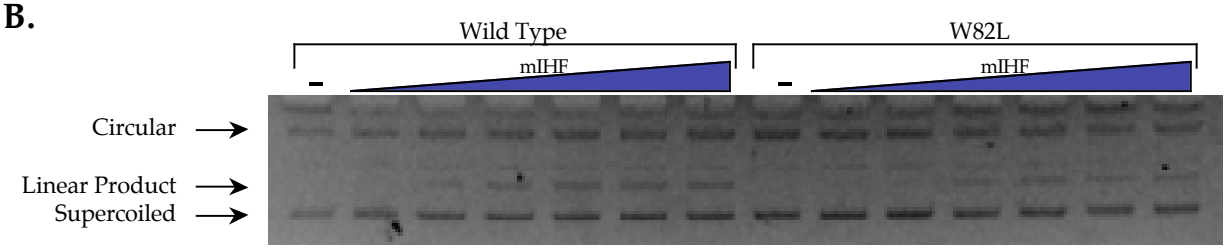
B. *In vitro recombination of pMK17 with increasing concentrations of mIHF*. To examine the relationship of mIHF to recombination efficiency of pMK17 containing reactions, *in vitro* recombination was carried out in the presence of increasing concentrations of mIHF (14 nM – 2.8 nM). 11.9 nM wild type Int or W82L was allowed to perform recombination for 2 hours at 37ºC under standard conditions and reactions were stopped by heating to 65ºC for 20 minutes. Products were separated on a 1% agarose gel and visualized by ethidium bromide staining under UV light. These results clearly show that both wild type Int and W82L are able to perform recombination on this substrate and that increasing the mIHF concentration increases the efficiency of this reaction.

C. *In vitro recombination of pMK17 with T92A, H138L and A150E and increasing concentrations of mIHF*. Reactions containing each of the middle domain Int mutants to determine if T92A, H138L, and A150E could perform recombination with pMK17 in the presence of high concentrations of mIHF. Reactions were carried out as described above. None of the mutants were able to carry out recombination at any of the concentrations tested.

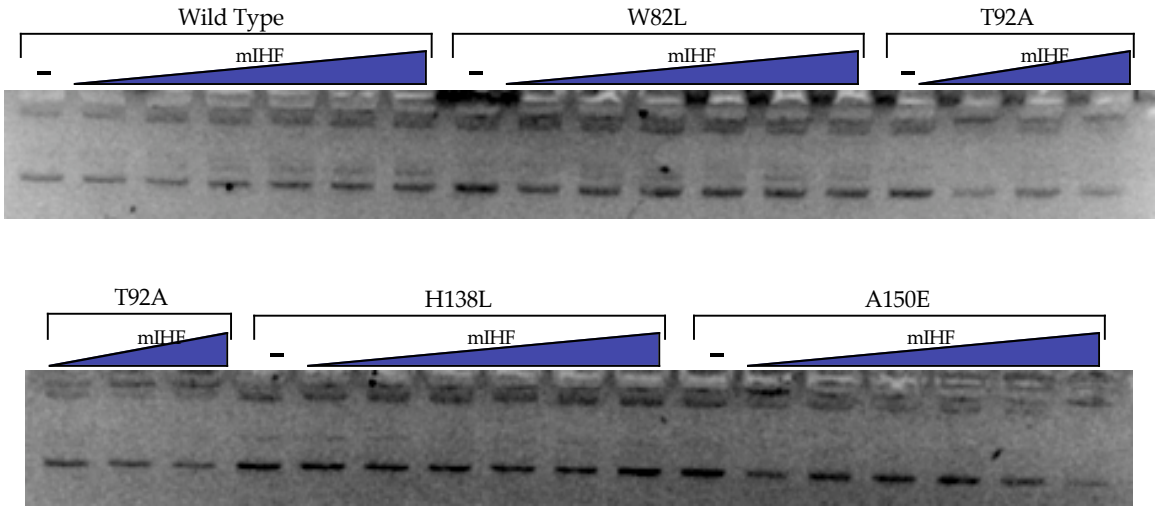
A.


Wild Type
W82L
5
15
30
60
90
120
5
15
30
60
90
120
Circular
Linear Product
Supercoiled

B.


Wild Type
W82L
mIHF
mIHF
Circular
Linear Product
Supercoiled

C.


Wild Type
mIHF
W82L
mIHF
T92A
mIHF
T92A
mIHF
H138L
mIHF
A150E
mIHF

recombination with *attP* substrates containing insertions of 4, 6, 9, 11, and 13 base pairs (pMK4, 13, 12, 6 and 17, respectively) under standard conditions and were incubated for 2 hours at 37°C. Both were able to perform recombination on substrates with 4, 9, 11, and 13 base pair insertions to varying degrees (Figure 34), but neither could perform recombination on pMK13, with a 6 base pair insertion between core and P4/P5. Therefore, pMK13 was used as substrate in reactions with a mIHF titration.

Reactions using pMK13 as substrate were carried out for all mutants as described above (Figure 35). After 2 hours of incubation at 37°C, no product was formed at any mIHF concentration for either wild type Int or any of the middle domain mutants. These results emphasize the importance of mIHF in forming a recombinationally active complex. Since increasing the mIHF concentration in reactions using *attP* insertion mutants that reduce the efficiency of recombination augments recombination, L5 Int may not be able to bring mIHF into the complex with substrates with longer linkers between core and arm-type binding sites. By increasing the mIHF concentration, the frequency with which mIHF finds the Int *attP* complex increases, and more active complex is formed enhancing the recombination efficiency. If middle domain mutants are less capable of bringing mIHF into the complex due to mutation in an area of the protein that forms protein/protein interactions between Int and mIHF, increasing the mIHF concentration will be less likely to rescue recombination and little or no product will be formed. This appears to be the case. Recombination is supported by pMK17 swith wild type Int and W82L, which appears to have the least severe phenotype based on recombination rates on various substrates and on the insertion mutant data. However, T92A, H138L and A150E are unable to perform recombination on the pMK17 *attP*.

Figure 34. pMK *attP* Insertion Mutants *in vitro* Recombination

Mutant *attPs* with insertions between P4/P5 and core pMK4 (+4 base pairs), pMK6 (+11 base pairs), pMK12 (+9 base pairs), pMK13 (+6 base pairs), pMK17 (+13 base pairs) and pCP□R13 (wild type) were used to test the ability of W82L to perform recombination on these substrates. *In vitro* reactions were set up under standard conditions with 11.9 nM wild type Int or W82L, approximately 3.4 nM each *attP* containing plasmid, 70 nM 42 base pair *attB*, and 280 nM mIHF. Recombination continued for 2 hours at 37°C before being heated to 65°C for 20 minutes to stop the reactions. Products were separated on a 1% agarose gel and visualized by ethidium bromide staining under UV light. Both proteins were able to perform recombination on all of the insertion mutant substrates except pMK13, which contains a 6 base pair insertion.

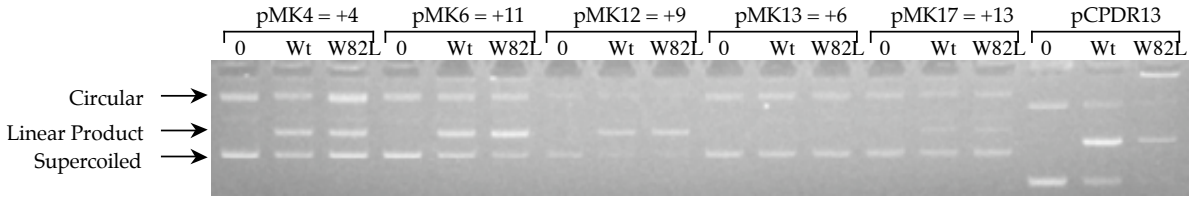
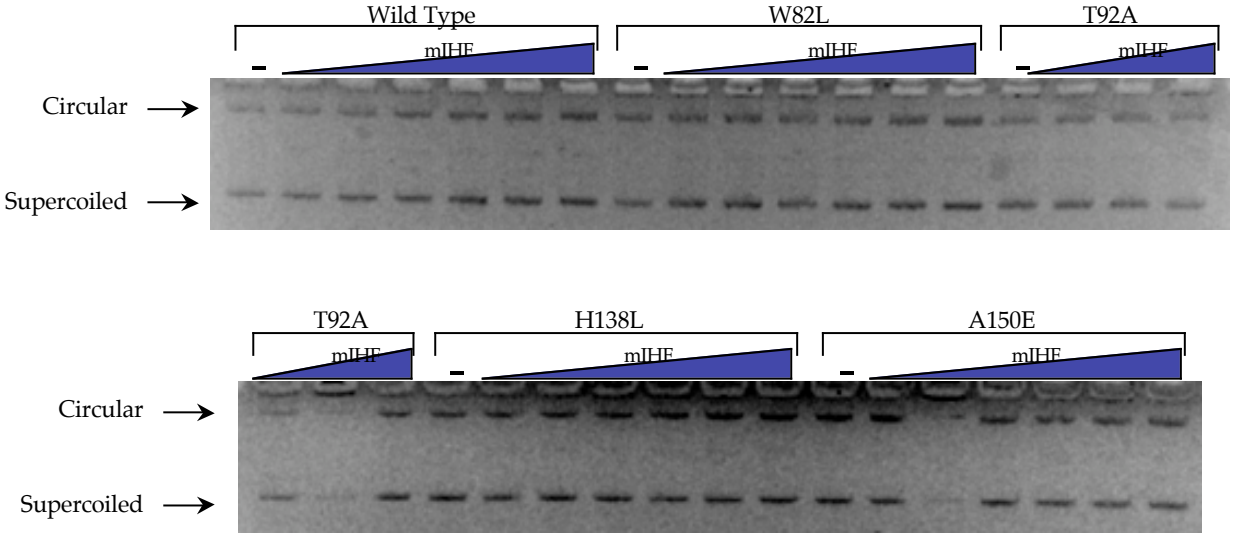

pMK4 = +4
pMK6 = +11
pMK12 = +9
pMK13 = +6
pMK17 = +13
pCPDR13
0 Wt W82L 0 Wt W82L 0 Wt W82L 0 Wt W82L 0 Wt W82L 0 Wt W82L
Circular
Linear Product
Supercoiled

Figure 35. pMK13 *in vitro* Recombination

In vitro recombination was carried out using 3.4 nM 7769 base pair pMK13, which contains a 6 base pair insertion between P4/P5 and core, as supercoiled substrate and increasing concentrations of mIHF. Reactions containing 11.9 nM Int, 70 nM 42 base pair *attB* and increasing concentrations of mIHF (14 nM to 2.8 μM) as well as pMK13 were allowed to perform recombination under standard conditions for 2 hours at 37°C. Products were separated using agarose gel electrophoresis on a 1% agarose gel and were visualized using ethidium bromide staining under UV light. The image has been inverted to show minute amounts of product. These data show that neither wild type Int nor any of the middle domain Int mutants can perform recombination on this substrate at any concentration of mIHF tested.


Wild Type
W82L
T92A
mIHF
mIHF
mIHF
Circular
Supercoiled
T92A
H138L
A150E
mIHF
mIHF
mIHF
Circular
Supercoiled

IV.I. Intasome is Stabilized at Increased mIHF Concentrations

Initial EMSA data suggests that W82L, T92A and H138L are capable of forming weak intasome in the presence of wild type *attP* and mIHF. However, none of these mutants are able to form synaptic complex 1 or 2 when in the presence of *attB* (Figure 27). Mutant A150E was unable to form any distinct complexes either in the presence or absence of *attB*. Because mIHF concentration affects the ability of these mutants to perform recombination on linear and supercoiled substrates and the ability of wild type and W82L to perform recombination on *attP* spacing mutants, mIHF may also help stabilize recombination complexes in the middle domain mutants.

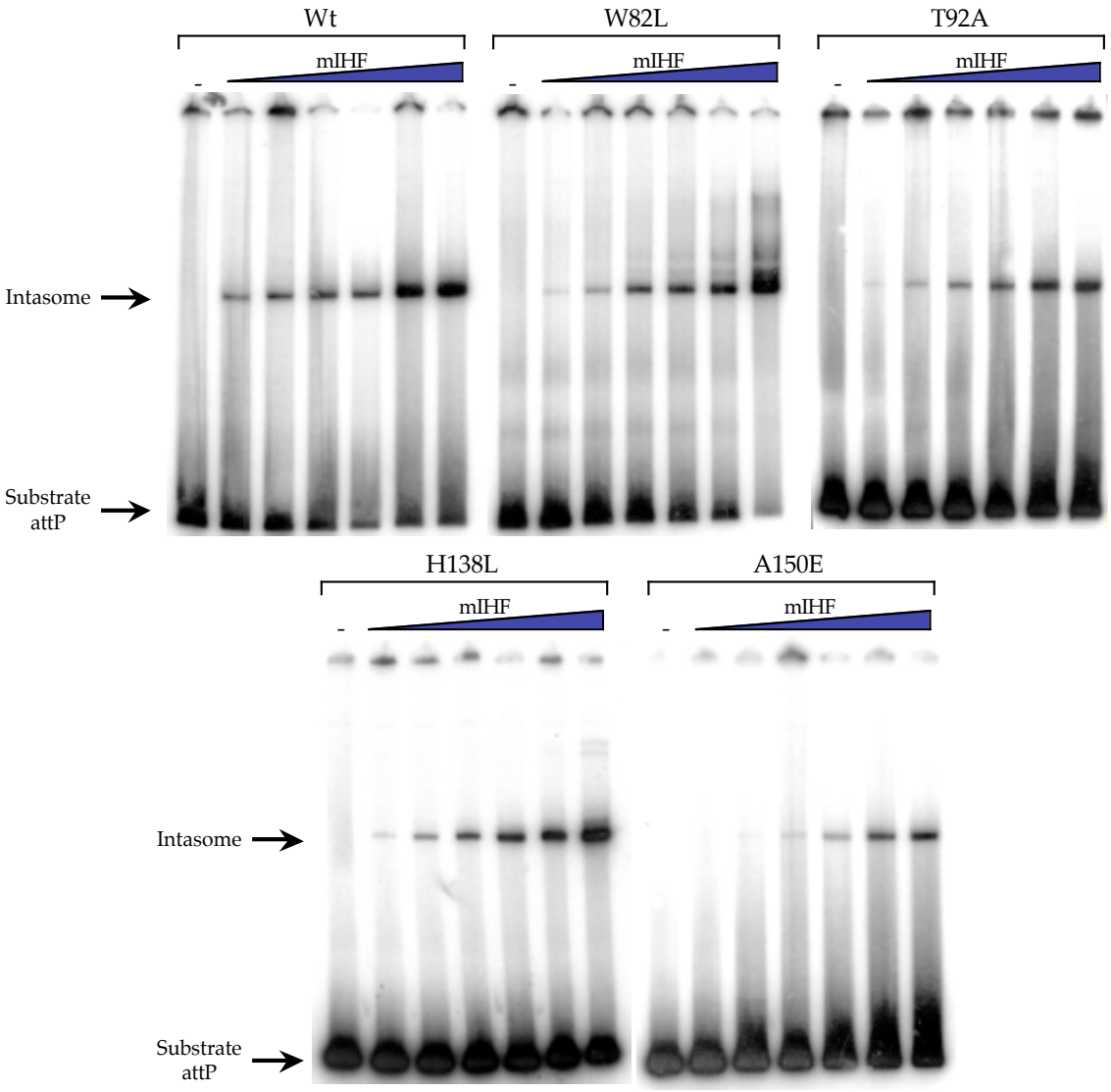
To test this hypothesis, band shift assays for intasome formation were carried out by mixing 11.9 nM Int, wild type and mutants W82L, T92A, H138L and A150E, with radio labeled 342 base pair *attP* and increasing concentrations of mIHF (14nM - 5.6 μM). Complexes were allowed to form on ice for 20 minutes before being separated by 5% native poly-acrylamide gel electrophoresis and visualized by autoradiography (Figure 36).

Wild type Int forms intasome well at even the lowest concentration of mIHF (14 nM), and intasome formation is increased as a function of mIHF concentration. Weak intasome formation is observed in reactions containing W82L, T92A and H138L at low concentrations of mIHF. Intasome appears to be stabilized at high mIHF concentrations, like wild type. A150E intasome is also stabilized at high mIHF concentrations, but no intasome is observed until much higher mIHF concentrations indicating that this mutant has a more severe phenotype then the other mutants.

These results suggest that mIHF can stabilize intasome in middle domain Int mutants as well as wild type Int, since even mutant A150E, which appears to have the

Figure 36. Intasome Formation by Middle Domain Int Mutants

EMSA was used to test the ability of middle domain Int mutants and wild type Int to form intasome as a function of mIHF concentration. EMSA was performed by mixing 11.9 nM Int with increasing concentrations of mIHF (28 nM – 5.6 μM) and radio labeled *attP* under standard reaction conditions. Complexes were formed on ice for 20 minutes before being separated on a 5% non-denaturing poly-acrylamide gel run at 4ºC. Complexes were visualized by autoradiography. These data show more intasome is formed at higher concentrations of mIHF.


Wt
W82L
T92A
mIHF
-
Intasome
Substrate
attP
H138L
A150E

most severe phenotype, can form intasome at high mIHF concentrations. This stabilization could be due to the ability of mIHF to bend the *attP* DNA more readily at high concentrations. The stabilization may also be the result of more mIHF entering the *attP* Int complex and allowing the formation of higher quantities of less stable intasome. If there is direct contact between Int and mIHF, the resulting stabilization may be the effect of more mIHF contacting Int directly. Protein/protein interactions may trigger a conformational change in Int protomers causing their affinity for *attP* core to increase. Increasing the affinity of Int for *attP* core should not only stabilize intasome but should also stabilize *attB*/Int interactions.

IV.J. The Effect of *attB* Concentration on Recombination

Synaptic complexes 1 and 2 both contain *attB*. The inability of middle domain mutants, W82L, T92A, H138L and A150E, to form synaptic complexes suggests that these mutants interact with *attB* less effectively then wild type Int. If these mutants' apparent reduced affinity for *attB* affects their ability to perform recombination, increasing the *attB* concentration in *in vitro* reactions containing these mutants should enhance the ability of these mutants to perform recombination.

In vitro recombination was performed with both linear and supercoiled *attP* substrate. The concentration of both Int and mIHF were held constant, 5.95 nM and 280 nM, respectively, with *attB* being titrated into the reactions (93 nM -1.4 μM). Recombination was allowed to continue for 2 hours at 37°C before being stopped by heating to 65°C for 20 minutes. The resulting products were separated by agarose gel electrophoresis and visualized by ethidium bromide staining under UV light (Figure 37 A linear, Figure 38 A supercoiled). The product was quantified using BioRad Gel

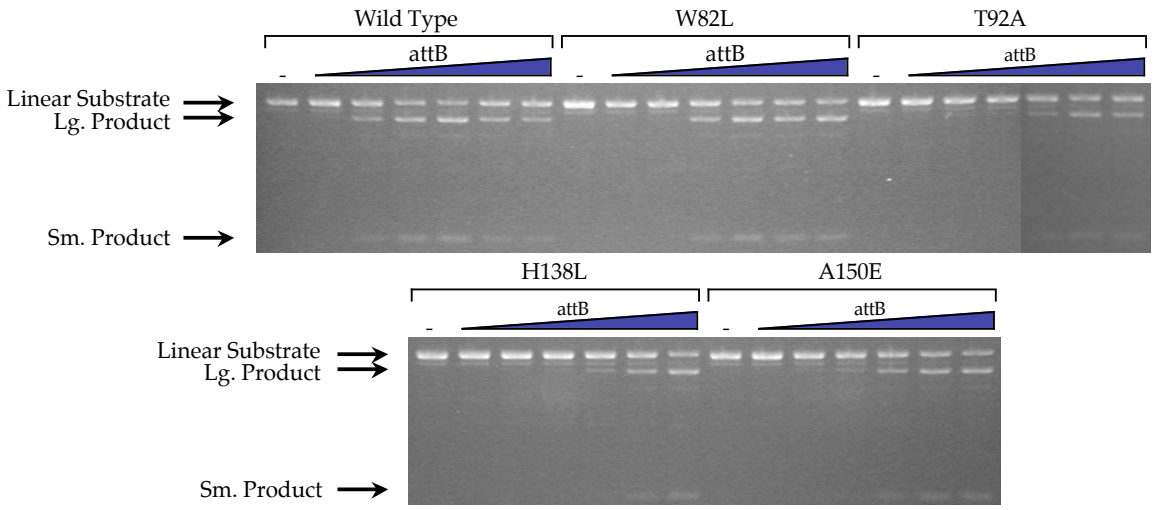
Figure 37. The Effect of *attB* Concentration on *in vitro* Recombination

The ability of wild type Int and middle domain Int mutants to perform recombination as a function of *attB* concentration was tested by performing *in vitro* recombination assays in the presence of varying concentrations of *attB* (46.5 nM – 700 nM). Assays were performed under standard conditions with 11.9 nM Int, 280 nM mIHF, 3.4 nM linear 5899 base pair *attP* and increasing concentrations of 42 base pair *attB*. Recombination occurred while the reactions were incubated for 2 hours at 37°C and was stopped by heating to 65°C for 20 minutes. Products were separated on a 1% agarose gel and visualized by ethidium bromide staining under UV light. Density of product bands was determined using BioRad Gel Documentation system and QuantityOne software.

A. *Recombination by middle domain and wild type Ints in the presence of increasing concentrations of attB.* The products of recombination are observed as a 5165 base pair band that runs just below the substrate band and a smaller 734 base pair fragment. The *attB* concentration for each set of reactions is increased from left to right. The efficiency of recombination by wild type Int is inhibited in the presence of high concentrations of *attB* as observed by a reduced intensity in the product band at high *attB* concentrations, however the middle domain mutants perform recombination with greater efficiency at high *attB* concentrations.

B. *Plot of the percent substrate converted to product versus the attB concentration.* The density of the large product band was determined, and the background for each lane was subtracted and divided by the total density of substrate. This value was multiplied by 100 to obtain the percent total DNA converted to product. The percent total product was then plotted versus the *attB* concentration.

A.


Wild Type
W82L
T92A
attB
attB
attB
-
-
-
Linear Substrate
Lg. Product
Sm. Product
H138L
A150E
attB
attB
-
-
Linear Substrate
Lg. Product
Sm. Product

B.

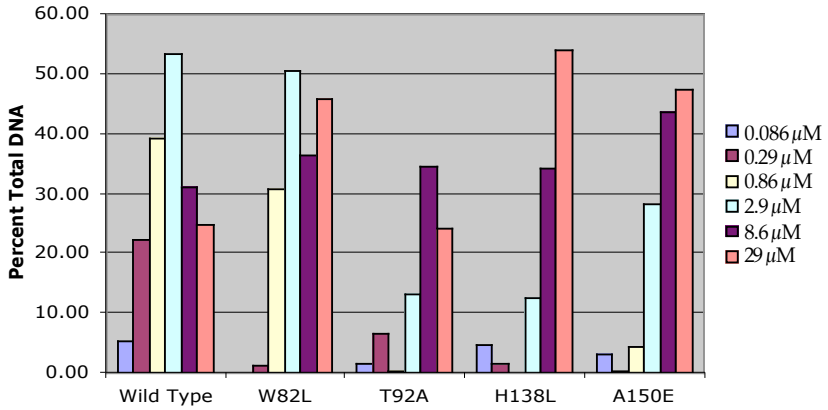

60.00
50.00
40.00
30.00
20.00
10.00
0.00
Percent Total DNA
Wild Type
W82L
T92A
H138L
A150E
0.086 μM
0.29 μM
0.86 μM
2.9 μM
8.6 μM
29 μM

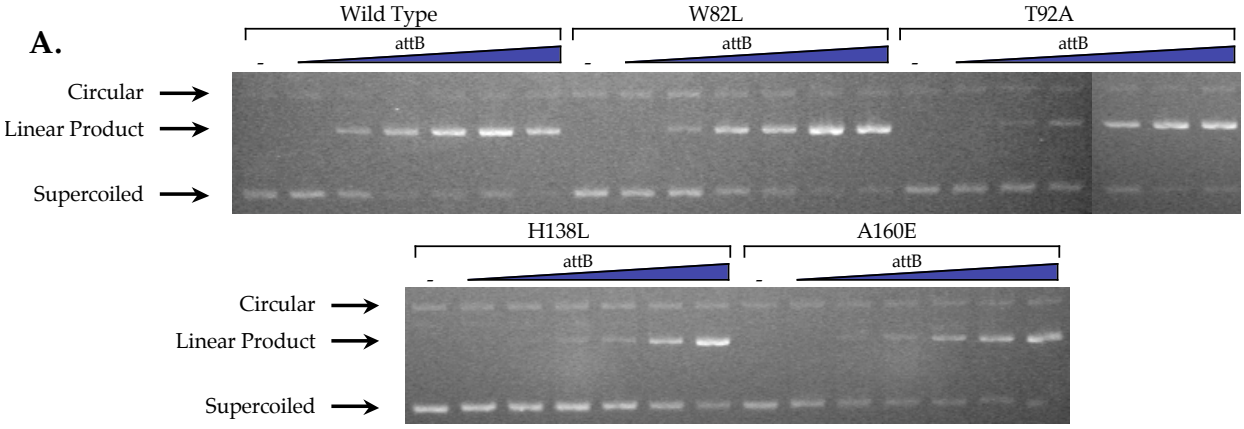
Figure 38. The Effect of *attB* Concentration on *in vitro* Recombination with Supercoiled Substrate

11.9 nM Int was used in recombination reactions under standard conditions containing 280nM mIHF, 3.4 nM supercoiled *attP*, and increasing concentrations of *attB* (46.5 nM – 700nM). Reactions were carried out for 2 hours at 37°C before being heat killed at 65°C for 20 minutes. Products were separated on a 1% agarose gel and visualized by ethidium bromide staining under UV light. Products were quantified using BioRad Gel Documentation system and QuantityOne software.

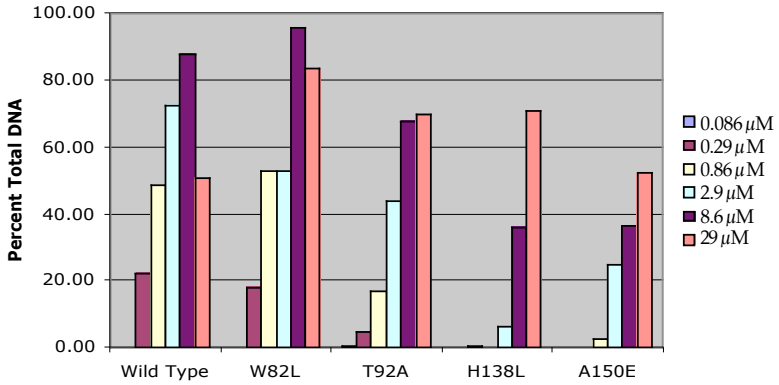
A. *In vitro recombination of supercoiled substrate with increasing concentrations of attB*. Super coiled substrate forms a linear product of recombination that runs between a lower supercoiled band and a higher nicked circular band. These data show that wild type Int is unable to perform recombination as efficiently at high concentrations of *attB*. Middle domain mutants perform recombination more efficiently at high *attB* concentrations.

B. *Plot of percent substrate converted to product versus the attB concentration*. The product band was quantified as explained above, and background for each lane was subtracted. The density of product was then divided by the total density of substrate and multiplied by 100 to get the percent total DNA converted to product. This value was then plotted versus the *attB* concentration.

A.


Wild Type
W82L
T92A
attB
attB
attB
Circular
Linear Product
Supercoiled
H138L
A160E
attB
attB
Circular
Linear Product
Supercoiled

B.


100.00
80.00
60.00
40.00
20.00
0.00
Percent Total DNA
Wild Type
W82L
T92A
H138L
A150E
0.086 μM
0.29 μM
0.86 μM
2.9 μM
8.6 μM
29 μM

Documentation system and QuanityOne software. These results were plotted versus mIHF concentration (Figure 37 B linear, Figure 38 B supercoiled).

These results suggest that the optimal *attB* concentration for wild type Int recombination is ~140 nM since more product is observed at this concentration then the others tested. Above or below this concentration, product formation is reduced for both linear and supercoiled attP substrates. Unlike wild type, W82L, H138L and A150E were all able to perform recombination more efficiently at higher concentrations of *attB*. T92A reacted much like wild type Int, but its optimal *attB* concentration appears to be higher at 930 nM. Above or below this concentration, product formation is reduced.

Middle domain Int mutants W82L, H138L and A150E perform recombination more effectively at high *attB* concentrations than wild type. This could reflect a lower affinity for *attB* in the middle domain mutants, since increasing the concentration of *attB* would at least partially compensate for the reduced affinity. The inhibition of recombination at high *attB* concentrations probably represents the formation of unproductive complexes as *attB* out competes *attP* core for binding to the catalytic domain. The reduced affinity of middle domain mutants for *attB* would restrict the formation of these unproductive complexes while increasing the ability of these mutants to form productive complexes at high *attB* concentration.

IV.K. Middle Domain Mutants are Unable to Form Synaptic Complex 2 at High *attB* Concentrations

Recombination efficiency appears to be affected directly by both mIHF and *attB* concentrations. mIHF has been shown to stabilize intasome at high concentrations, but do high concentrations of *attB* have the same effect? If the affinity of the middle domain Int mutants for *attB* is reduced, increasing the *attB* concentration should allow the

mutants to over come this reduced affinity and form synaptic complexes 1 and/or 2 at high *attB* concentrations.

Wild type Int has been shown to form synaptic complex 2 and a secondary complex containing Int forming an intermolecular bridge between P4/P5 and *attB* in the absence of mIHF (Pena et al. 2000). To test whether increasing the concentration of *attB* can stabilize weakened interactions between *attB* and the middle domain Int mutants, band shift analysis was performed in the absence of mIHF. Reactions containing radio labeled *attP*, 11.9 nM Int, and increasing concentrations of *attB* (93 nM – 1.4 μM) were created and complexes were formed on ice for 20 minutes. Complexes were separated by electrophoresis on a 5% non-denaturing poly-acrylamide gel and visualized by autoradiography (Figure 39).

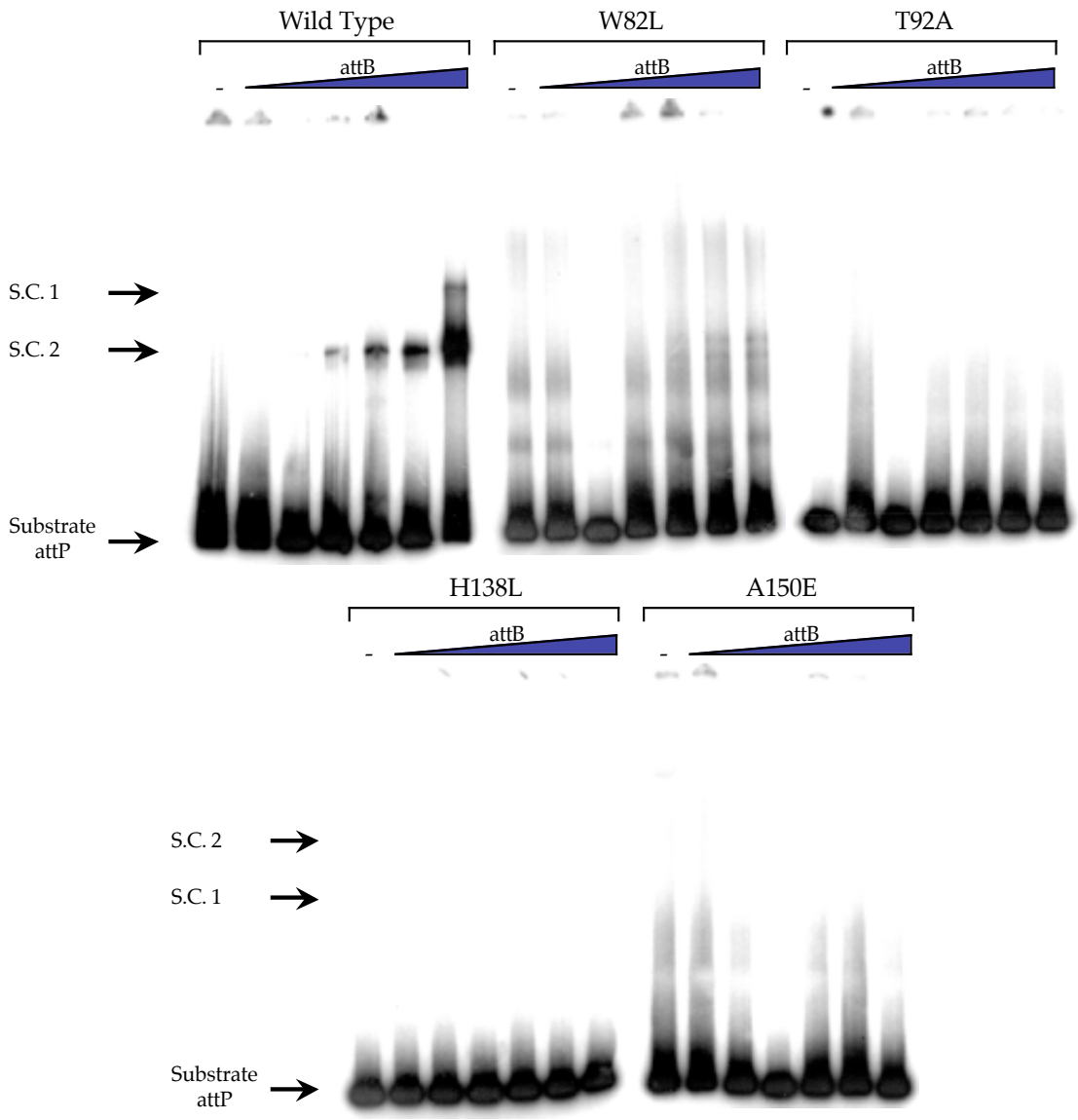
Wild type Int forms synaptic complex 2 at relatively low concentrations of *attB* (14nM), and at higher concentrations of *attB*, the secondary P4/P5 complex is formed. Mutants T92A, H138L and A150E do not form synaptic complex 1 at any *attB* concentration tested and never form the secondary P4/P5 complex. W82L forms a very weak complex that runs at approximately the same position as synaptic complex 1 as well as a second band that runs just below this position. Although this complex may be synaptic complex 2, very little is produced at 100 times the concentration of *attB* necessary to see this complex in wild type Int. These results suggest that high *attB* concentrations are unable to compensate for the apparent lower affinity of middle domain mutants for *attB*.

IV.L. W82L Forms Synaptic Complex 1 to at High mIHF but not *attB* Concentrations

Both mIHF and *attB* are able to increase the efficiency of recombination for the middle domain mutants. High concentrations of mIHF are able to stabilize intasome.

Figure 39. Synaptic Complex 2 Formation of Middle Domain Int Mutants

To test the ability of middle domain mutants to form complexes with *attB*, EMSA was performed in the absence of mIHF. Each reaction contained 11.9 nM Int, increasing concentrations of *attB* (93 nM – 1.4 μM), and radio labeled 342 base pair *attP*. Complexes were allowed to form on ice for 20 minutes before being separated by electrophoresis on a 5% non-denaturing gel run at 4°C. Autoradiography was used to visualize complexes. Wild type Int forms synaptic complex 2 at relatively low concentrations of *attB* and the P4/P5 *attB* bridging complex at higher *attB* concentrations. W82L may form synaptic complex 2 at very high *attB* concentrations, but none of the other mutants are able to make this complex at any of the *attB* concentrations tested.


Wild Type
W82L
T92A
attB
-
S.C. 1
S.C. 2
Substrate
attP
H138L
A150E

If increasing the mIHF concentration stabilizes intasome, recombinationally active complexes may be stabilized in the same way allowing recombination to occur more efficiently at high mIHF concentrations. This does not appear to be the case for *attB*. Increasing the *attB* concentration in *in vitro* reactions allows more efficient recombination, but increasing the *attB* concentration effects complex formation very little in the absence of mIHF indicating that *attB* and mIHF interact with each other.

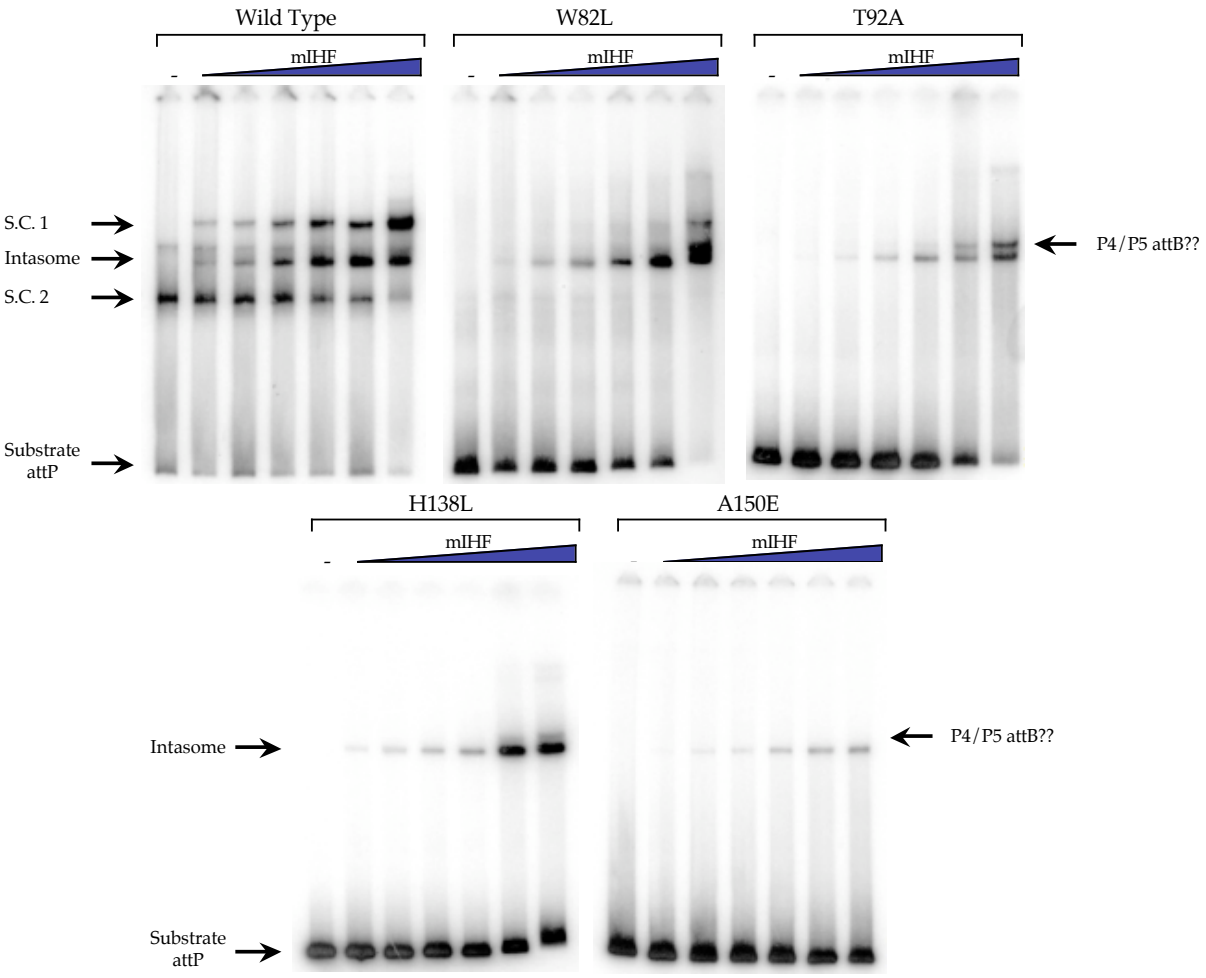
To test the ability of the middle domain mutants to form synaptic complexes, EMSA was performed by mixing 11.9 nM of each Int, radio labeled 342 base pair *attP*, and increasing concentrations of mIHF (14 nM - 5.6 μM) in the presence of 144 nM of *attB*. Complexes were allowed to form on ice for 20 minutes and were separated by electrophoresis on a 5% non-denaturing poly-acrylamide gel (Figure 40).

In the absence of mIHF, wild type Int forms synaptic complexes 2 and a P4/P5 *attB* intermolecular bridging complex that runs just above the intasome. As mIHF is added to the reaction, synaptic complex 1 becomes visible as synaptic complex 2 and the secondary P4/P5 complexes are lost. Very little synaptic complex 2 remains at high mIHF concentrations as synaptic complex 1 becomes more stable. The amount of intasome is also lowered at high mIHF concentrations. Mutants T92A, H138L and A150E are unable to form either synaptic complex at any mIHF concentration tested. However at very high mIHF concentrations (5.6 μM), a complex that runs at the same position as the secondary P4/P5 complex is formed by mutant T92A and weakly by H138L. W82L forms this complex, as well as a complex that runs at a similar position to synaptic complex 2.

These results suggest that high concentrations of mIHF are able to stabilize intasome in the middle domain mutants. Intasome also appears to be stabilized in the presence of *attB* and high mIHF concentrations, but no synaptic complexes are formed

Figure 40. Synaptic Complex Formation at High mIHF Concentrations

EMSA was performed with 11.9 nM Int in the presence of 140 nM *attB*, increasing concentrations of mIHF (28 nM – 5.6 μM) added to reaction sets from left to right, and radio-labeled *attP* under standard conditions. Complexes were formed on ice for 20 minutes and separated on a 5% non-denaturing gel run at 4ºC. Wild type Int forms synaptic complex 2 in the absence of mIHF. As mIHF is added, intasome and synaptic complexes 1 are formed. Middle domain mutants form intasome at all concentrations of mIHF tested, but are unable to form either synaptic complex. W82L appears to form synaptic complex 1 at high mIHF concentrations.


Wild Type
W82L
T92A
mIHF
-
S.C. 1
Intasome
S.C. 2
Substrate attP
P4/P5 attB??
H138L
A150E
Intasome
Substrate attP
P4/P5 attB??

with mutants T92A, H138L or A150E. H138L, T92A and W82L form a complex that runs just above intasome, but it would seem unlikely that this is actually the secondary P4/P5 complex since no synaptic complex 1 is formed and it has been shown that Intermolecular bridging between P1/P2 and *attB* are preferred to intermolecular bridges between P4/P5 and *attB* (Peña and Hatfull 2000). W82L forms a complex that runs at a similar position to synaptic complex 2. However, the composition of this complex cannot be discerned from these experiments.

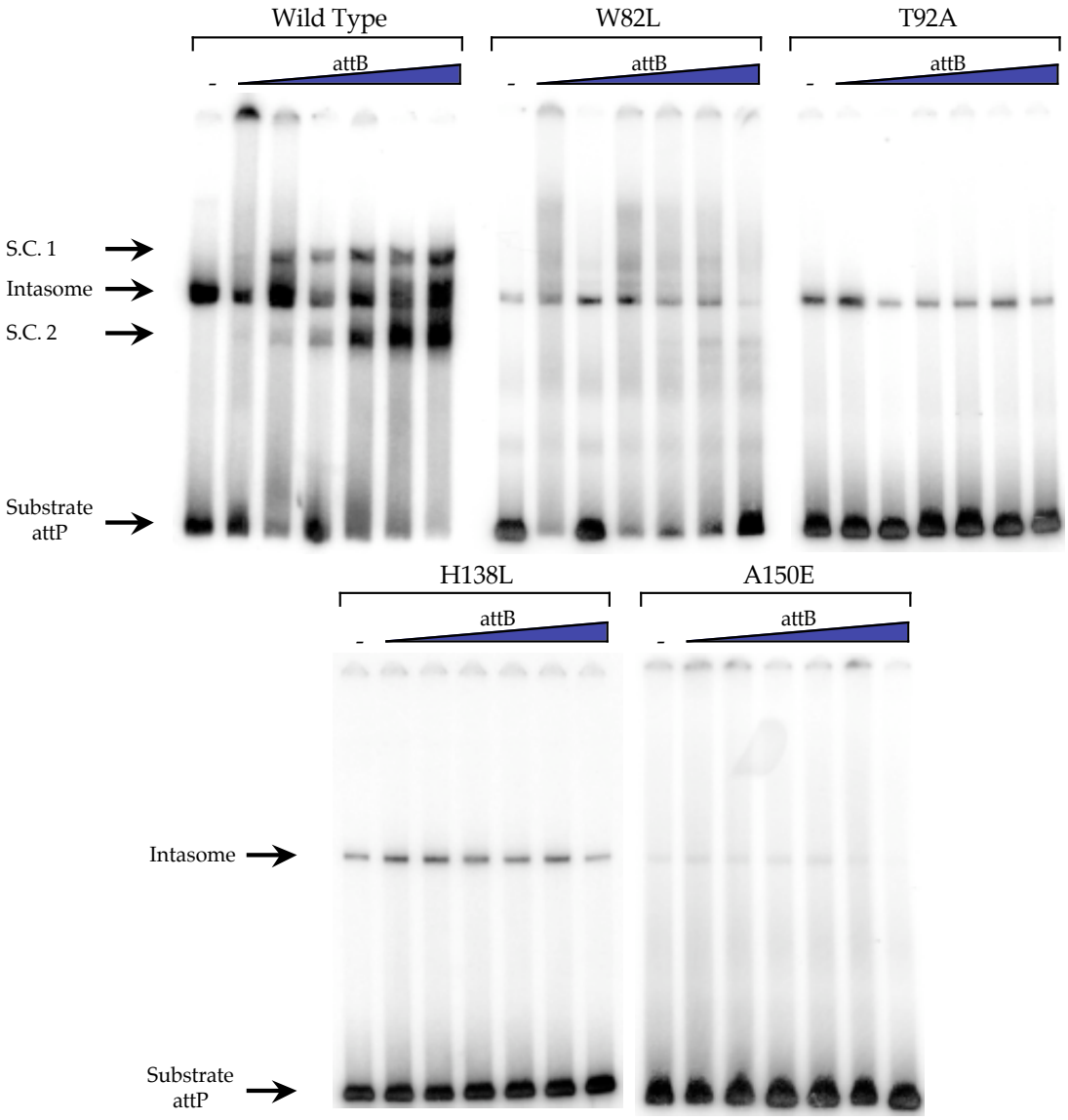
EMSA was carried out in the presence of increasing concentrations of *attB* to test the ability of *attB* to stabilize synaptic complexes in middle domain mutants. Complexes were formed on ice for 20 minutes in reactions containing 11.9 nM Int, wild type and each of the middle domain Int mutants, 560 nM mIHF, radio-labeled *attP* and increasing concentrations of *attB* (93 nM – 1.4 μM). The resulting complexes were separated by native gel electrophoresis on a 5% non-denaturing poly-acrylamide gel and visualized using autoradiography (Figure 41).

In the absence of *attB*, intasome is formed in reactions containing wild type Int. As the *attB* concentration is increased, formation of synaptic complexes 1 and 2 can be observed. The secondary P4/P5 *attB* bridging complex is also observed at high *attB* concentrations. No synaptic complexes are formed in reactions containing any of the middle domain mutants, W82L, T92A, H138L, and A150E, even at the highest *attB* concentration (1.4 μM). These mutants show intasome formation at comparable levels to those in previously described experiments, but this complex may become destabilized at high *attB* concentrations as witnessed by a decrease in intasome intensity in the final lanes of W82L and A150E sets.

These results along with the inability of these mutants to form synaptic complex 2 in the absence of mIHF indicate that *attB* alone is unable to stabilize any of the

Figure 41. Synaptic Complex Formation at High *attB* Concentrations

EMSA was performed in the presence of 11.9 nM Int, 560 nM mIHF, increasing concentrations of *attB* (93 nM – 1.4 μM) and radio-labeled 342 base pair *attP* under standard reaction conditions. Complexes were formed on ice for 20 minutes and separated by electrophoresis on a native 5% poly-acrylamide gel. Complexes formed were visualized using autoradiography. In the absence of *attB*, wild type Int forms intasome. When *attB* is added, synaptic complexes 1 and 2 are formed. None of the middle domain mutants are able to form synaptic complexes at any *attB* concentrations, and intasome may be destabilized at high *attB* concentrations.


Wild Type
W82L
T92A
attB
-
S.C. 1
Intasome
S.C. 2
Substrate
attP
H138L
A150E
Intasome
Substrate
attP

complexes formed by Int and *attP*. However, mIHF retains the ability to stabilize intasome and an unknown secondary complex and possibly synaptic complex 1 in reactions containing W82L.

To further examine the ability of mIHF to stabilize synaptic complexes and to look more closely at *attB*'s relationship to this stabilization, EMSA was used to observe complexes formed at both high mIHF (Figure 42 A) and high *attB* (Figure 42 B) concentrations. Complexes containing 10 times the normal concentrations of mIHF or *attB* in the presence of increasing concentrations of *attB* or mIHF, respectively, were allowed to form on ice and separated by native gel electrophoresis as described above. Complexes were visualized by autoradiography.

When high concentrations of *attB* (1.4 μM) are present in EMSA reactions containing wild type Int synaptic complex 2 and the P4/P5 *attB* complex are formed in the absence of mIHF. As mIHF is added, intasome and synaptic complex 1 are also formed. It should be noted that synaptic complex 2 is not lost as readily as at lower concentrations of *attB*. None of the middle domain Int mutants form a synaptic complex 2 in the absence of *attB* as observed above. As mIHF is added, intasome is formed similarly to previously described experiments. The secondary unknown complex that runs above intasome created by W82L and H138L is observed and W82L appears to form synaptic complex 1 at high mIHF concentrations.

When the opposite experiment is performed, high mIHF with increasing *attB* concentrations, more synaptic complex 2 is observed. In the absence of *attB*, wild type Int forms intasome at high mIHF concentrations, as do all of the middle domain mutants. As *attB* is titrated into these reactions, synaptic complexes 1 and 2 are formed by wild type Int. T92A, H138L and A150E form stable intasome at all concentrations of *attB* but do not form synaptic complexes. H138L forms the secondary complex that

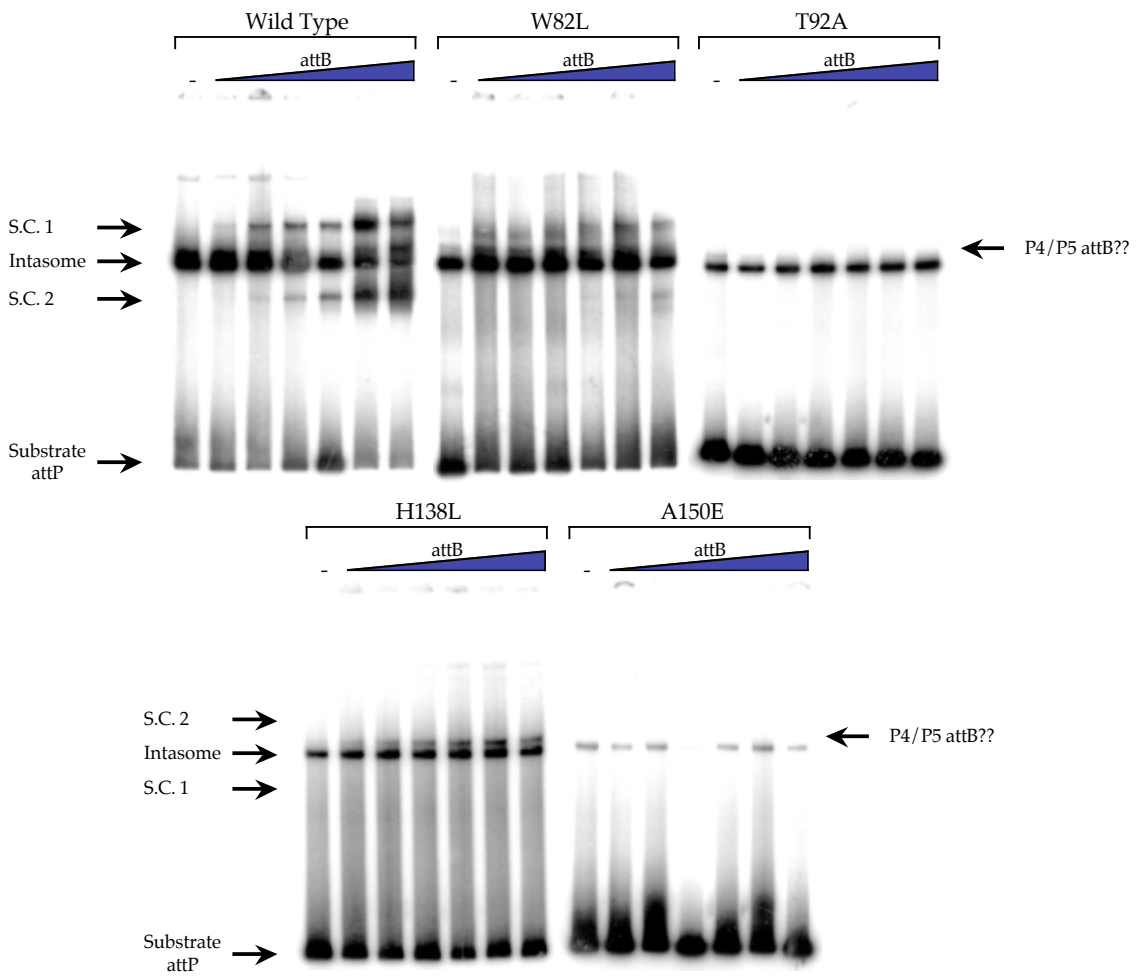
Figure 42. Synaptic Complex Formation at High mIHF and *attB* Concentrations

EMSA was used in the presence of high mIHF or high *attB* concentrations to determine the relationship between these factors.

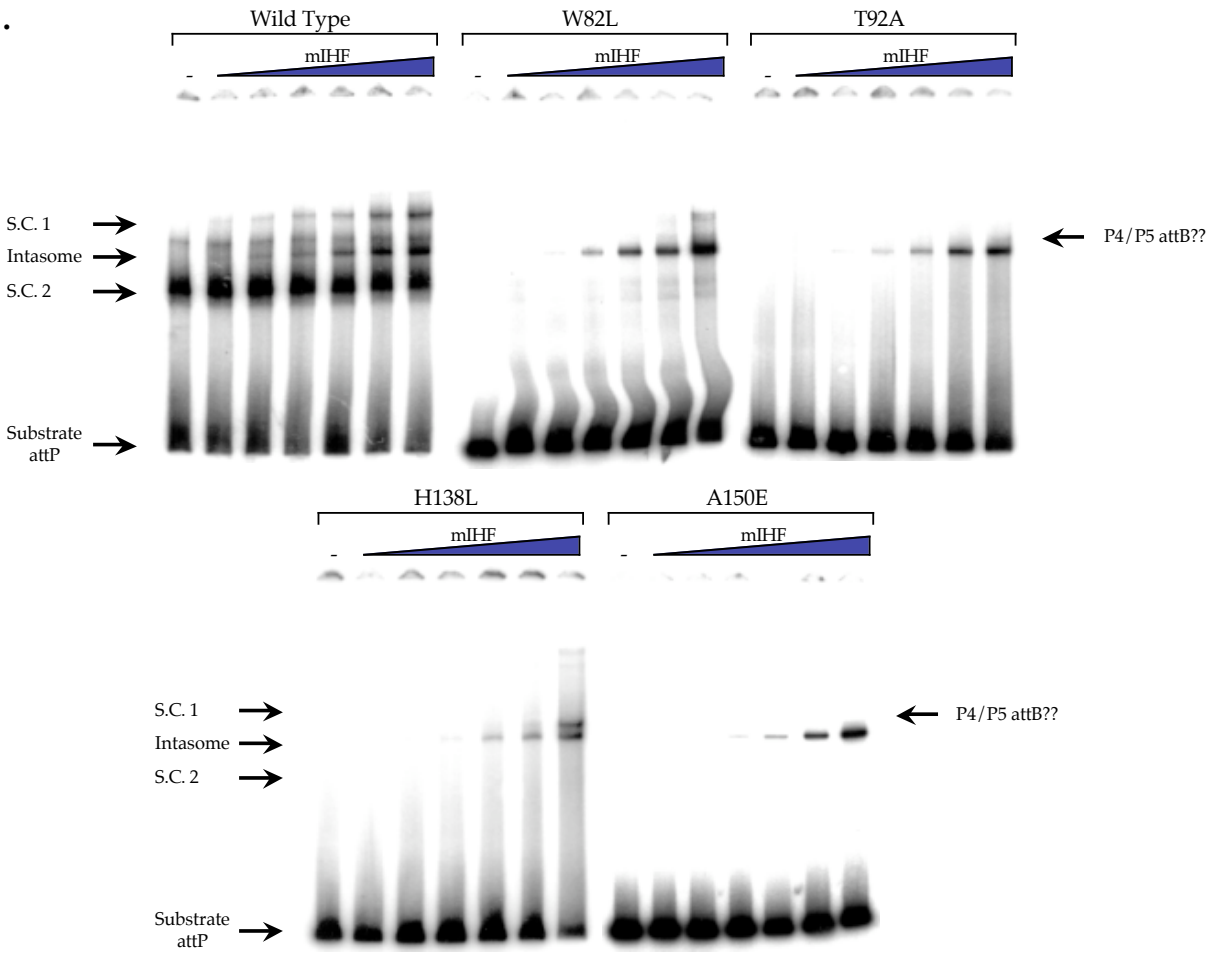
A. *Complex formation at high mIHF and increasing attB concentrations*. EMSA was performed at 10 times the standard concentration of mIHF (5.6 μM) in the presence of 11.9 nM Int, radio labeled *attP*, and increasing concentrations of *attB*. Complexes were formed under standard conditions on ice for 20 minutes and were separated on a 5% non-denaturing gel. Complexes were visualized by autoradiography. Wild type Int forms intasome in the absence of *attB* and synaptic complexes 1 and 2 as *attB* is added to the reaction. W82L appears to form synaptic complex 1 at 144 nM *attB* and above. None of the other mutants are able to form either synaptic complex.

B. *Complex formation at high attB and increasing mIHF concentrations*. In the presence of high concentrations of *attB* (1.4 μM), with increasing concentrations of mIHF (28 nM – 5.6 μM) under the conditions described above, wild type Int forms strong synaptic complex 2 and Intasome and synaptic complex 1 as the mIHF concentration is increased. W82L forms synaptic complex 1 only at high concentrations of mIHF. The other middle domain mutants form intasome, but neither synaptic complex at any of the mIHF concentrations tested.

A.


Wild Type
W82L
T92A
attB
-
S.C. 1
Intasome
S.C. 2
Substrate
attP
P4/P5 attB??
H138L
A150E
attB
-
S.C. 2
Intasome
S.C. 1
Substrate
attP
P4/P5 attB??

B.


Wild Type
W82L
T92A
mIHF
-
S.C. 1
Intasome
S.C. 2
Substrate
attP
P4/P5 attB??
H138L
A150E

runs above intasome. The intensity of this complex increases with the *attB* concentration indicating that it contains *attB*. W82L again forms a complex that runs in the same position as synaptic complex 1. This complex also appears to be responsive to *attB* concentration and forms at *attB* concentrations above 144 nM *attB*.

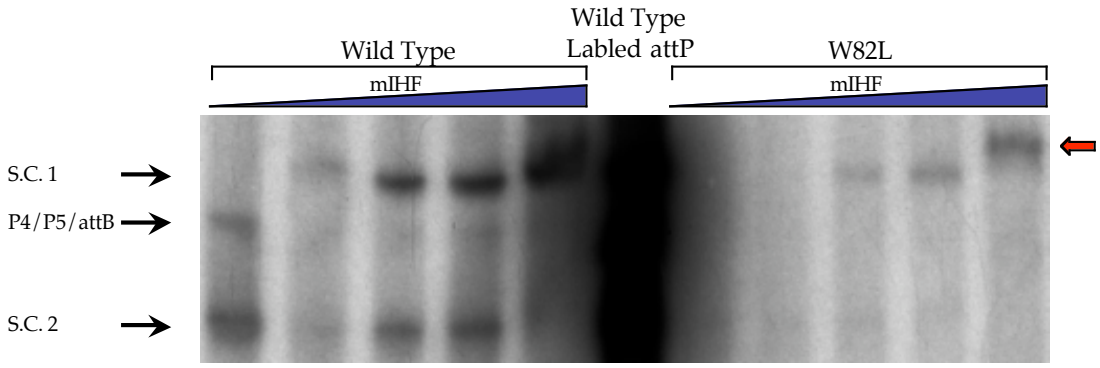
These results suggest that mIHF is acting to stabilize the Int *attB* interaction. Taken together with the stabilization of intasome at high concentrations of mIHF, this could mean that mIHF is acting to stabilize the Int *attP* core interaction. The ability of W82L to form more stable intasome than the other mutants and to form synaptic complexes indicates that this mutant has a less severe phenotype than the other mutants, so these mutations represent varying degrees of the same phenotypes with A150E having the most severe and T92A and H138L having intermediate phenotypes. Since H138L is able to form a secondary complex that runs at approximately the same location as the P4/P5 *attB* complex seen in wild type Int, this mutant may be deficient in forming a P1/P2 *attB* intermolecular bridge, but mIHF is allowing the stabilization of the P4/P5 *attB* intermolecular bridge.

IV.M. W82L Forms Synaptic Complex 2 at High mIHF Concentrations

To ensure that the complex formed by W82L at high mIHF concentrations is synaptic complex 2, EMSA was carried out using radio-labeled *attB*. These reactions contained approximately the same concentrations of *attP* and *attB* as previously described experiments, 11.9 nM Int, wild type and each of the middle domain Int mutants, and increasing concentrations of mIHF. Complexes were allowed form on ice for 40 minutes and were separated using native gel electrophoresis. The complexes formed were visualized by autoradiography (Figure 43).

Figure 43. Synaptic Complex 1 and 2 Formation at High *attB* and mIHF Concentrations

To ensure that the complex formed by W82L is synaptic complex 1, EMSA was performed at 11.9 nM Int, 3.4 nM unlabeled 342 base pair *attP*, increasing concentrations of mIHF (28 nM- 5.6 μM) and radio labeled *attB*. Complexes were formed on ice for 20 minutes and separated by electrophoresis on a 5% non-denaturing poly-acrylamide gel. Results were visualized by autoradiography. These results clearly show that W82L is able to capture *attB* at high mIHF concentrations.


Wild Type
Wild Type
Labled attP
W82L
mIHF
mIHF
S.C. 1
P4/P5/attB
S.C. 2

Wild type Int forms two complexes that contain *attB*, synaptic complex 2 that is formed first and migrates below intasome followed by synaptic complex 1 which runs higher on the gel above intasome. W82L is the only middle domain mutant that forms complex with *attB*. This complex runs at the same position as synaptic complex 2 and is only formed at high mIHF concentrations. These results show that the complex formed by W82L is synaptic complex 2. This supports the previously described results suggesting that mIHF is stabilizing not only the Int/core *attP* interaction observed through increased intasome formation, but also stabilizes the Int/*attB* interaction.

IV.N. The Catalytic Domain Mutant Tyrosine 342 to Phenylalanine Forms Wild Type Complexes

The middle domain mutants described above are unable to form synaptic complexes 1 and 2 and perform recombination less efficiently then wild type Int. To determine whether the inability to form synaptic complexes is indicative of middle domain mutants or a side effect of reduced recombination efficiency, the catalytic tyrosine of wild type Int was changed to phenylalanine (Y342F) creating a catalytically inactive mutant. Y342F was then subjected to in vitro recombination and EMSA, and its phenotype was described in terms its ability to perform recombination and form complexes.

Site-directed mutagenesis was carried out on pPSC1, a pET21a derived expression vector containing the L5 *Int* gene, using the Stragene site-directed mutagenesis kit. This created a mutant Int that contained a single amino acid substitution, a phenylalanine for a tyrosine, at position 342. Following mutagenesis, the entire *Int* gene was sequenced to ensure that no other mutations occurred during the mutagenesis process. The change from tyrosine to phenylalanine should create an Int

mutant that is catalytically inactive without perturbing the over all structure of the protein. Therefore, Y342F should form the same complexes like wild type Int but be unable to perform recombination.

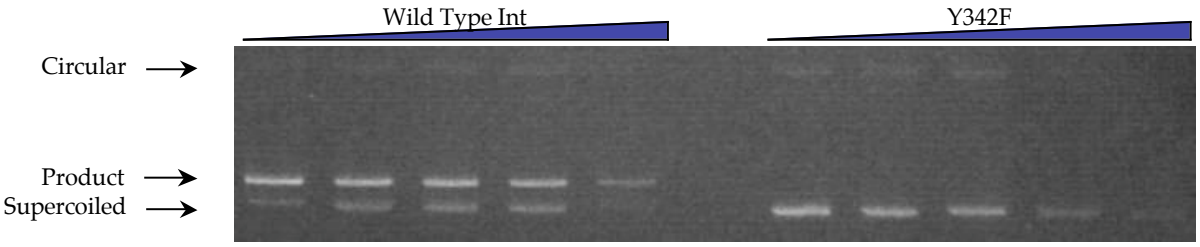
In vitro recombination was used to show that Y342F is catalytically inactive. Reactions containing 4 nM supercoiled *attP* containing pCP□R13, 280 nM mIHF, 70 nM *attB* and increasing concentrations of Y342F were incubated for 2 hours at room temperature to allow recombination to occur. These reactions were then stopped by heating to 65ºC for 20 minutes. Products were separated by electrophoresis on a 1% agarose gel and visualized by ethidium bromide staining under UV light (Figure 44). *In vitro* recombination was performed in parallel using wild type Int to compare product formation. This was also carried out using radio-labeled *attB* so that even minute amounts of product could be able to be distinguished.

In the presence of even low concentrations of wild type Int, linear product was observed running between a faster moving supercoiled substrate band and a slower moving nicked circular band. No product was observed at any of the Y342F concentrations tested. At high Int concentrations, substrate is bound and not released after heating. These products run in the wells or as smears on the gel and are observed as a disappearance of substrate. This is noted for both wild type Int and Y342F indicating that Y342F is still able to bind *attP*.

The ability of Y342F to interact within the recombinagenic complex was then tested by *in vitro* competition. *In vitro* recombination was carried out as described above with 11.9 nM wild type Int. Y342F was titrated into these reactions to observe the effect of inactive mutants on wild type recombination. Recombination was allowed to take place for 2 hours at room temperature before being stopped by heat killing. The

Figure 44. Y342F *in vitro* Recombination

In vitro recombination was performed under standard conditions using 4 nM supercoiled *attP* containing pCP□R13, 280 nM mIHF, 70 nM *attB* and increasing concentrations of wild type Int or Y342F for 20 hours at room temperature. Recombination was stopped by heating to 65ºC for 20 minute and visualized by ethidium bromide staining under UV light. Wild type Int forms linear product that runs between faster moving supercoiled substrate and slower moving nicked circular substrate at all Int concentrations tested. Y342 is unable to perform recombination at any Y342F concentration, but is able to bind *attP*-containing substrate at high protein concentrations retaining the substrate in the wells like wild type Int.


Wild Type Int
Y342F
Circular
Product
Supercoiled

products were separated on a 1% agarose gel and visualized by ethidium bromide staining under UV light (Figure 45).

In the absence of Y342F, Int forms linear product well as visualized by the dense band between lower supercoiled and higher running nicked circular substrate. However, as Y342F is added to the reactions, less product formation is observed. This is probably due to Y342F forming competing for *attP* binding. Since Y342F inhibits recombination at low concentrations (0.4 nM), the inactive mutant is probably able to complex with wild type Int at the synapse. Recombination is blocked since one or more of the protomers in the recombinagenic complex are unable to perform the nucleophilic attack on the substrate DNA. At high Y342F concentrations, no product is made. These data show that mutation of the catalytic tyrosine does not affect Ints ability to bind *attP* or form protein/protein intreactions with other Int protomers.

Y342F is catalytically inactive, but appears to bind *attP* and interact with other Int protomers like wild type Int. EMSA was used to show Y342F forms complexes like wild type Int independently. EMSA was performed by mixing 11.9 nM Int with 560 nM mIHF, 140 nM *attB*, and radio-labeled *attP* under standard reaction conditions and allowing complexes to form on ice for 20 minutes before being shifted to room temperature for increasing amounts of time (15 – 120 minutes). Complexes formed were separated by 5% native PAGE and visualized by autoradiography (Figure 46).

When Int is incubated with *attP*, mIHF, and *attB*, wild type Int forms intasome, synaptic complex 2, and synaptic complex 1. The distribution of these complexes changes as reactions are incubated at room temperature for increasing amounts of time as *attL* and *attR* product bands appear. Y342F forms all of the complexes described above but does not produce *attL* and *attR* product bands.

Figure 45. Wild Type Int/Y342F Competition *in vitro*

In vitro recombination was carried out in the presence of 11.9 nM Int, 4 nM supercoiled *attP* containing pCP□R13, 280 nM mIHF, 70 nM *attB* and increasing concentrations of Y342F. In the absence of Y342F, Int forms a linear product band that runs between faster moving supercoiled substrate and slower moving nicked circular substrate. As Y342F is titrated into the reactions, less product is observed indicating that Y342 is interfering with wild type Ints ability to perform recombination.

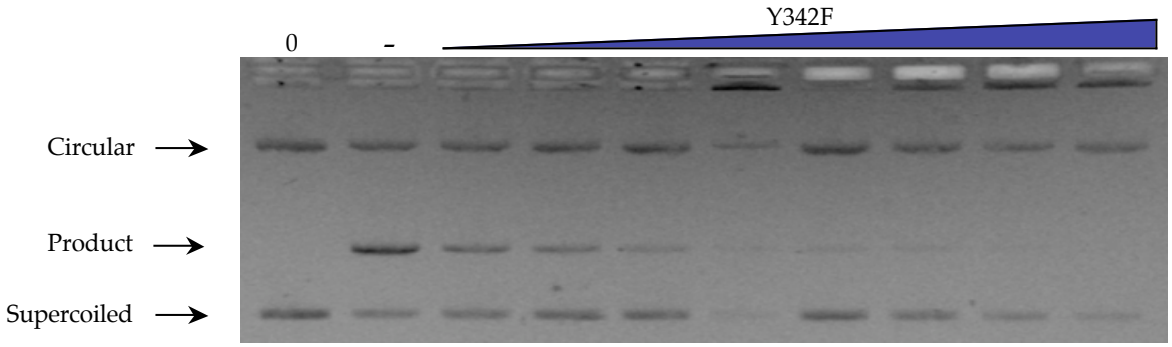
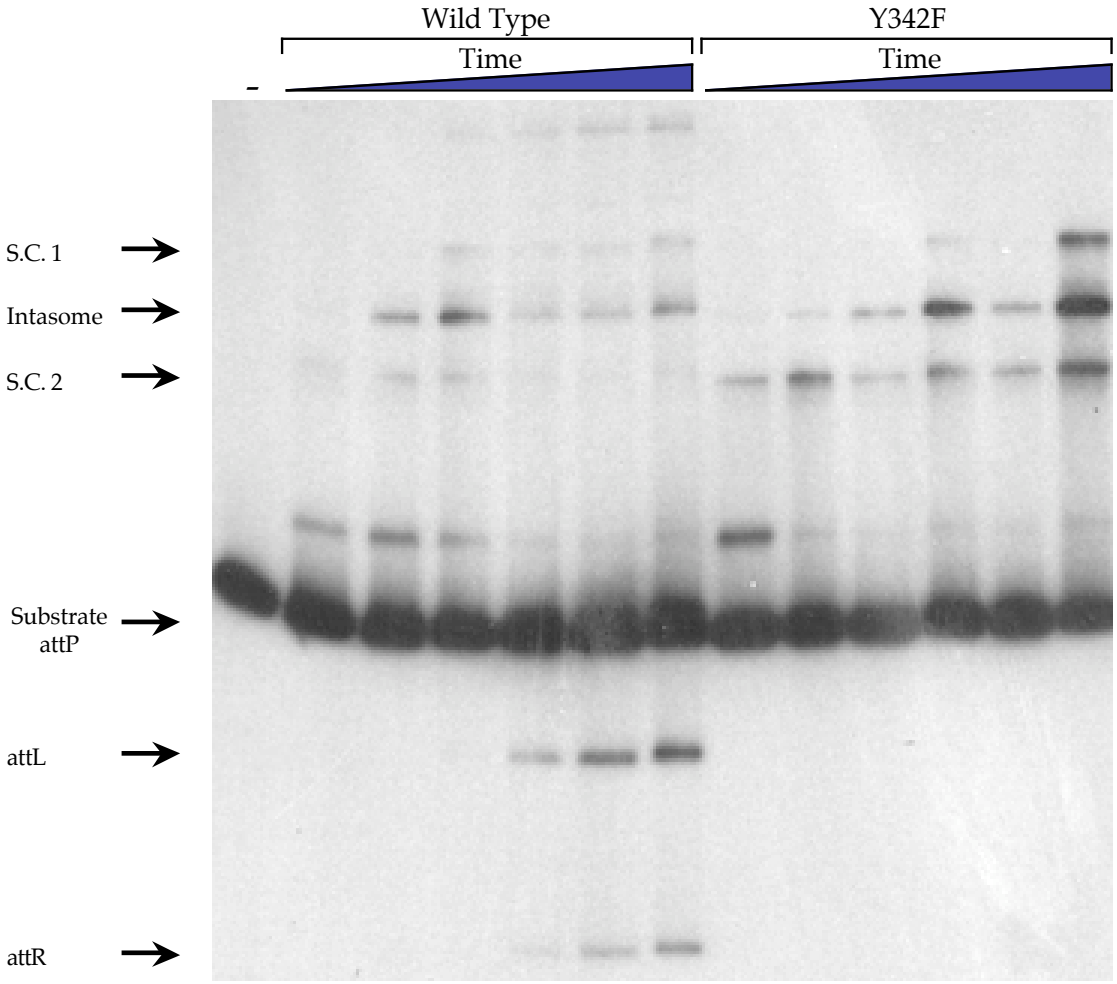

Y342F
0
-
Circular
Product
Supercoiled

Figure 46. Complex Formation of Y342F on Ice and at Room Temperature

EMSA was used to test the ability of Y342F to form complexes on ice and at room temperature. 11.9 nM wild type Int or Y342 was mixed with 560 nM mIHF, 140 nM *attB*, and radio-labeled *attP* under standard reaction conditions for at least 20 minutes on ice before being moved to room temperature for increasing amounts of time (5 – 120 minutes). Complexes were separated on a 5% poly-acrylamide gel and visualized by autoradiography. Wild type Int forms intasome as well as synaptic complexes 1 and 2 on ice and at room temperature. Products (*attL* and *attR*) are formed after incubation at room temperature and run below the substrate band. Y342F forms complexes like wild type Int on ice and at room temperature, but does not form product after incubation at room temperature for any amount of time tested.


Wild Type
Y342F
Time
Time
-
S.C. 1
Intasome
S.C. 2
Substrate
attP
attL
attR

The production of stable complexes by a catalytically disabled mutant shows that complex formation is not indicative of productive recombination, but rather shows the ability of Int to bind to the components of the recombination complex stably. Y342F appears to be able to form all of the intercomponent interactions between Int protomers, Int and mIHF, and Int and *attP* and *attB*. Since middle domain Int mutants are unable to form synaptic complexes, these data suggest that middle domain Int mutants are defective for interactions with one or more of the components in the recombinagenic complex.

IV.O. High Concentrations of mIHF are able to Increase the Efficiency of Recombination on Substrates with Mutant RBEs

Previous data has shown that Int is able to bind *attP* core with mutations in the left side RBE and catalyze recombination on this substrate, but when right side RBE is mutated, recombination is inhibited, and neither synaptic complex 2 nor intasome are not formed (Chapter III, Section F). It was also shown that *attP* mutants with mutations in both RBEs unable to be bound by Int or utilized as substrate in recombination (Chapter III, Section E). If mIHF is able to stabilize the Int/core interaction, high concentrations of mIHF may be able to stabilize the interaction between Int and the mutant *attP* cores allowing recombination to take place on substrates with RBE mutations.

In vitro recombination was carried out using pCWCM1, pCWCM2, pCWCML.5, and pCWCMR.5 in the presence of increasing concentrations of mIHF to examine the ability of mIHF to stimulate recombination of these mutant *attP* cores. Reactions containing 11.9 nM Int, 280 nM *attB*, 70 nM supercoiled pCWCM1, pCWCM2, pCWCML.5, or pCWCMR.5, and mIHF (14 nM to 2.8 μM) incubated at 37^{o}C for 2 hours

to allow recombination before being stopped by heating to 65ºC for 20 minutes. Products were separated on a 1% agarose gel and visualized by ethidium bromide staining under UV light (Figure 47).

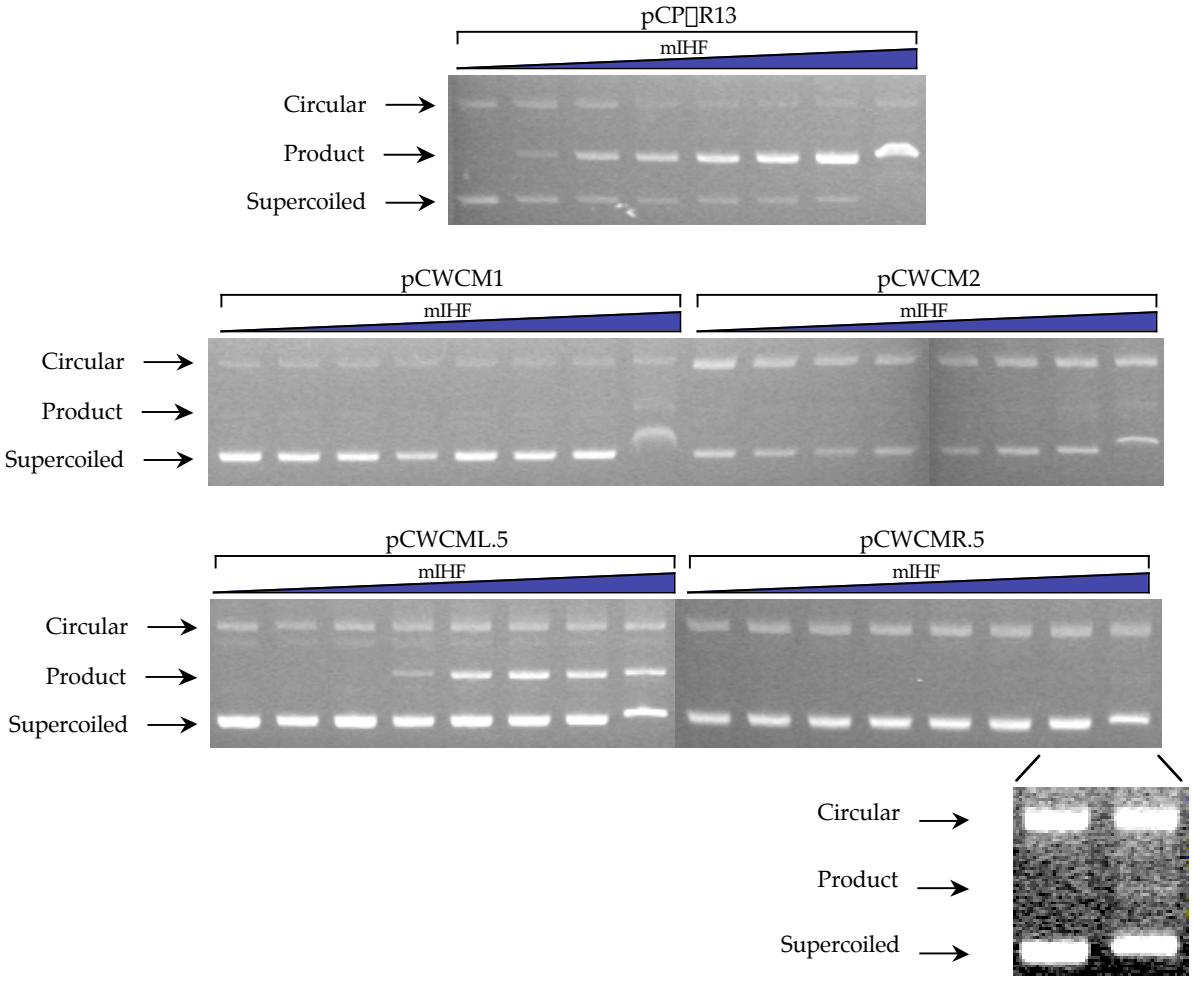
Int forms linear product that runs between faster moving supercoiled substrate and slower moving nicked circular substrate. As the mIHF concentration is increased, the density of the visible product increases to ~560 nM where product formation reaches it's peak. Recombination is eventually inhibited at high mIHF concentrations, as previously described. Recombination is not supported pCWCM1 and pCWCM2 at any mIHF concentration. Left side RBE mutant pCWCML.5 supports recombination at low levels in the presence of low concentrations of mIHF. As the concentration of mIHF is increased, the intensity of the product band increases, indicating that mIHF is stabilizing the Int/*attP* core interaction with the mutant substrate. The 7 base pair overlap region of the right side RBE has been mutated in pCWCMR.5. This mutant *attP* is unable to support recombination under standard conditions, but recombination is stimulated at very high mIHF concentrations (2.8 μM). This shows that the right side RBE is required for Int mediated recombination indicating that this RBE is recognized first. The Int/core interaction is stabilized by mIHF in both mutants, but pCWCMR.5 requires extremely high concentrations of mIHF for this effect.

IV.P. W82L Shows Loss of Coopertivity Based on *attP* Binding Affinity

Kinetic and EMSA data suggest that Int binds *attP* more tightly in the presence of mIHF since the Int/core interaction is stabilized by mIHF. To examine this relationship more closely, band shifts containing increasing concentrations of Int (0 – 2.4 μM) and radio-labeled *attP* were quantified. Band shift analysis was carried out as previously described. Components were mixed under standard conditions and

Figure 47. Wild Type Int *in vitro* Recombination on pCWCM1, pCWCM2, pCWCML.5, and pCWCMR.5 as a Function of mIHF Concentration

Reactions containing 11.9 nM Int, 280 nM mIHF, 70 nM *attB*, and mIHF (14 nM to 2.8 μM) incubated at 37°C for 2 hours to allow recombination and reactions were stopped by heating to 65°C for 20 minutes. Products were separated on a 1% agarose gel and visualized by ethidium bromide staining under UV light. Wild type Int forms linear product that runs between supercoiled substrate (lower band) and nicked circular substrate (higher band) at all mIHF concentrations tested when wild type *attP* containing pCP□R13 is provided as supercoiled substrate. Neither pCWCM1 nor pCWCM2 are able to support recombination at any mIHF concentration. pCWCML.5 is used as substrate more readily in the presence of high concentrations of mIHF. The same may be true for pCWCMR.5 although minute amounts of product are produced at 2.8 μM. The 0.93 and 2.8 μM lanes have been enlarged here to show the product in the 2.8 μM lane.


pCP□R13
mIHF
Circular
Product
Supercoiled
pCWCM1
mIHF
pCWCM2
mIHF
Circular
Product
Supercoiled
pCWCML.5
mIHF
pCWCMR.5
mIHF
Circular
Product
Supercoiled
Circular
Product
Supercoiled

reactions were incubated on ice for 20 minutes. Complexes were separated by electrophoresis on 5% native poly-acrylamide gel, and visualized by autoradiography (Figure 48 A). The density of the substrate band was then determined using a Fuji Phosphorimager system with Image Gauge software, and plotted versus the concentration of Int. These data were fit with a hyperbola (Equation 1) and a value for Equation 1.

$$y = \frac{\max(x)}{K_d + x}$$

the K_A of Int to full length *attP* was derived (Figure 48 A and B) Quantative band shifts were carried out in the absence of mIHF and at 3, 12 and 33 nM mIHF and values of K_A were determined at each concentration (Table 4). The resulting K_A values for the Int *attP* interaction were then plotted versus the mIHF concentration (Figure 49).

Wild type Int binds *attP* relatively weakly with an affinity of approximately $2x10^6$ M^{-1}. The addition of mIHF does not significantly change the affinity of Int for *attP* at low concentrations. However, the complex is reorganized as noted by the appearance of intasome. The formation of intasome without an increase in affinity for *attP* could indicate that the addition of mIHF results in the loss of unproductive interactions between the arm-type binding region of *attP* and the N-terminal domain of Int. At high concentrations of mIHF (>33 nM), the affinity of Int for *attP* is almost doubled. This appears to be a threshold affect, indicating a cooperative interaction between two binding sites on the same Int molecule or tetramer. Presumibly, if the concentration of mIHF was increased further, the affinity would again level off, creating a sigmoidal plot.

Middle domain mutants T92A, H138L, and A150E, do not show significant differences from wild type affinities for *attP*, however W82L may interact with *attP*

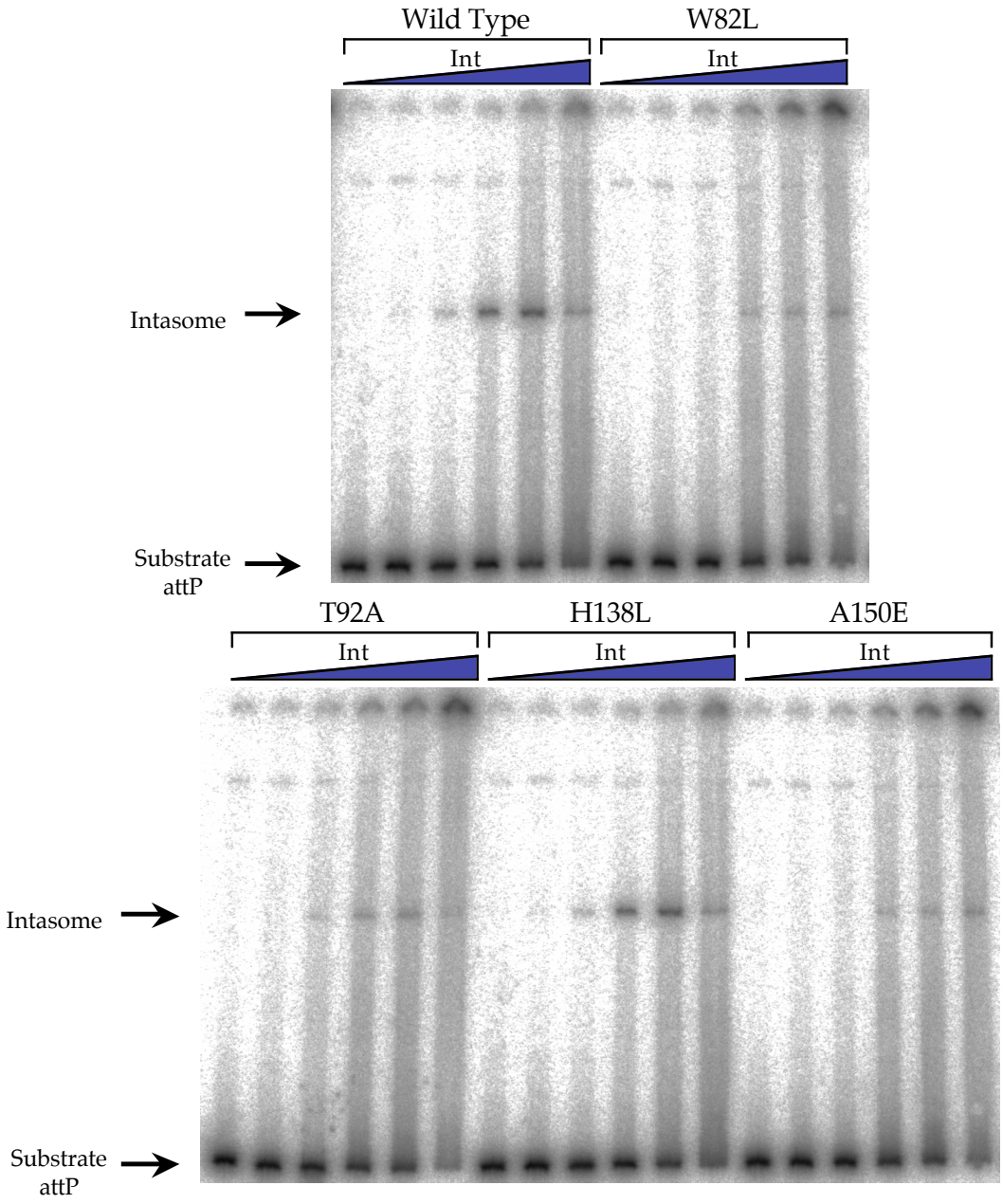
Figure 48. Representative Data From Quantitative EMSA

EMSA was carried out using increasing concentrations of Int (72 nM -2.4 μM) with radio-labeled *attP* under standard conditions and increasing concentrations of mIHF (0 – 33 nM). Complexes were formed on ice for 20 minutes and were separated on a 5% non-denaturing poly-acrylamide gel. Results were visualized by autoradiography.

A. *Representative data from EMSA performed at 3 nM mIHF*. The concentration of Int is increased from left to right, and as the Int concentration increases the amount of *attP* in the substrate band decreases indicating that Int is binding to the *attP* containing substrate. Complexes formed run in the well, as intasome, or as smears.

B. *Plot of percent substrate lost versus Int concentration*. The density of the substrate band was determined using a Fuji Phosphorimager with ImageGauge software. The substrate density was subtracted from the density the DNA added to each reaction and divided by total DNA. The resultant was then multiplied by 100 to give the percent substrate lost. This percentage was then plotted versus the Int concentration (shown here). Values for K_A were determined at each concentration of mIHF using ProFit 6.0 that fit these data to a hyperbola.

A.


Wild Type
W82L
Int
Int
Intasome
Substrate
attP
T92A
H138L
A150E
Int
Int
Int
Intasome
Substrate
attP

B.

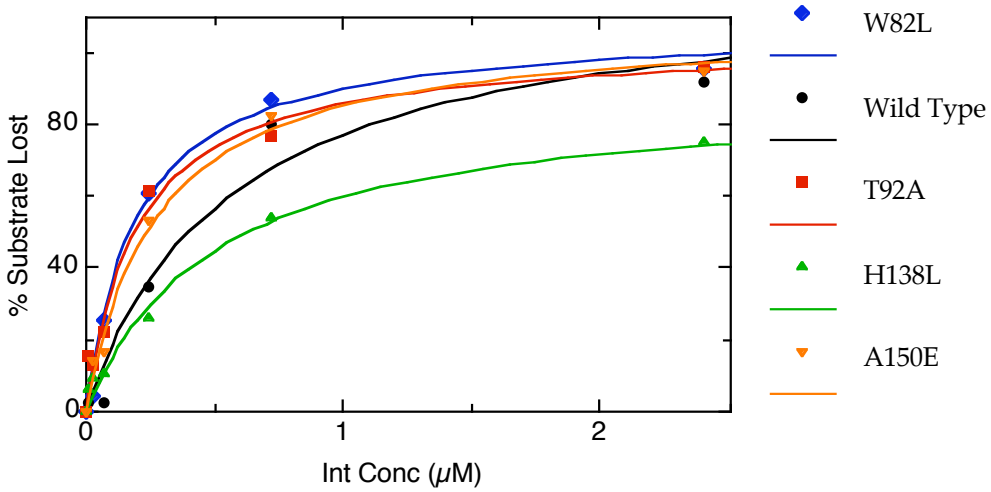

% Substrate Lost
0
40
80
0
1
2
Int Conc (μM)
W82L
Wild Type
T92A
H138L
A150E

Table 4. Binding Affinity of Middle Domain Int for *attP*

MIHF Concentration (μM)	K_A Wild Type	K_A W82L	K_A T92A	K_A H138L	K_A A150E
0.000	2.00±0.70*	3.30±0.21	3.42±0.72	0.73±0.14	1.71±0.31
0.003	1.91±0.83	4.42±0.56	3.43±0.86	1.85±0.62	2.61±0.91
0.012	1.66±0.99	4.12±1.44	6.61±1.76	1.31±0.56	5.38±1.73
0.033	4.37±1.09	4.34±0.92	5.11±1.08	2.33±0.50	2.93±0.71

*Binding constant times 10^6 ± standard deviation

differently. W82L appears to bind *attP* with a higher affinity then wild type Int in the absence of mIHF, and its affinity for *attP* is more similar of the affinity of wild type Int for *attP* at high (33 nM) mIHF concentrations. This could indicate that the W82L mutation causes a loss of coopertivity between binding sites, however the range of concentrations tested must be extended to fully observe this affect. It has been shown that Int binds arm-type *attP* DNA with higher affinity than core (Kim and Landy 1992, Minter et al. 1986), and this is believed to be the case for L5 Int. The W82L mutation is in the area of the protein were the middle and N-terminal domains converge, and may enhance the ability of the N-terminal domain to bind arm-type binding regions of *attP* by locking the N-terminal and middle domains into a conformation that allows tighter binding to arm-type sites.

IV.Q. Discussion

Recombination, as performed by L5 Int, is a complicated process involving recombinagenic protein/DNA complex that contains several protomers of Int binding *attP* and *attB*, as well as a host factor, mIHF. Int builds this complex by first binding to *attP* presumably at arm-type binding sites, P4/P5 (Kim and Landy 1992, Peña and Hatfull 2000). Int then recruits mIHF, which bends *attP* bringing *attP* core into position for binding to Int's catalytic domain.

Conformational changes within the catalytic domain of Int protomers are almost certainly a part of recombinagenic complex building process. Structural data from other tyrosine recombinases that do not require host factors suggest that binding of substrates causes a conformational change within the integrase that holds the substrates in place during recombination (Van Duyne 2001, Van Duyne 2002). Int may perform similar comformational changes when substrates are bound, but binding of a host factor

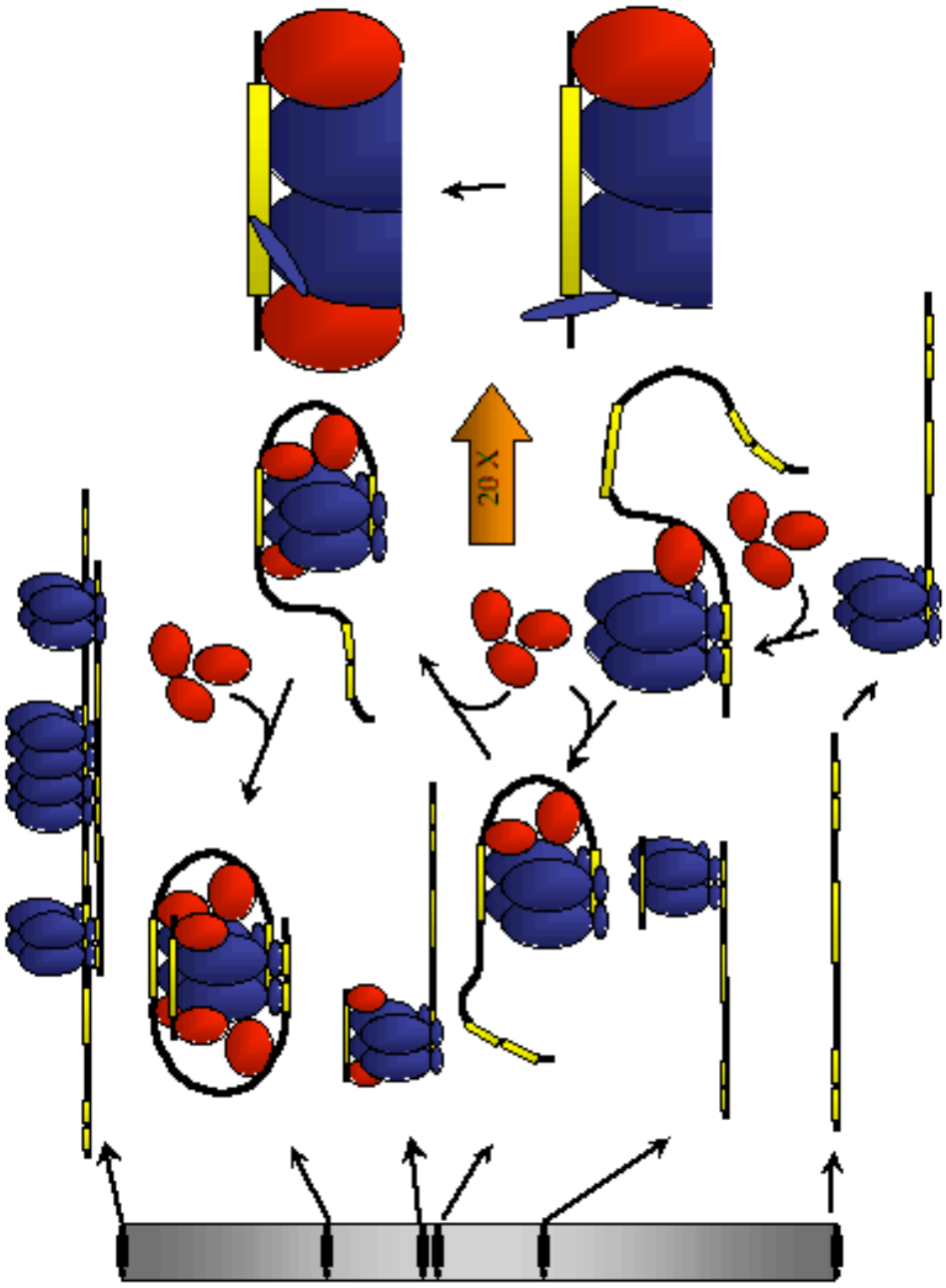
to the recombinagenic complex could also catalyze conformational changes that allow Int to bind *attP* more tightly (Figure 49).

Host factors involved in recombination, like mIHF, are known to catalyze *attP* bending that is required to bring *attP* core into position for Int binding (Better et al. 1982, Kim and Landy 1992, Rice et al. 1996, Richet et al. 1988, Thompson and Landy 1988). Binding sites for these host factors have been mapped to regions on *attP* between core and arm-type binding sites (Nash 1977, Nash 1981), but little is known about how these factors interact with Int protomers themselves, if they interact directly with Int at all. In at least one case, the binding site for IHF has been mutated, but IHF still catalyzes the bending of *attP* in the proper location through "pseudo-specific" binding to DNA in this region (Segall et al. 1994). DNA binding by mIHF is not site-specific, but *attP* is contacted at particular locations between the arm and core binding regions of *attP* (Pedulla and Hatfull 1998). DNA topology, or protein/protein interactions between Int and mIHF could help position mIHF on *attP*.

There are many cooperative interactions between Int protomers that allow the recombinagenic complexes to form properly. The most C-terminal helix (helix N) is donated to neighboring promomers, and in some cases, this helix contains catalytic residues (Gou et al. 1999). There are also a number of protein/protein interactions that occur in the N-terminal region of the catalytic domain (Gou et al. 1997). The N-terminal arm-type binding has been shown to form cooperative interactions with arm-type binding sites on *attP* (Sarkar, Azaro et al. In press). Reports also suggest that Xis forms direct interactions with Int protomers, and these interactions may be able to stabilize Ints binding to arm-type binding sites on lambda *attR* (Thompson et al. 1987b). The number and type of protein/protein interactions that occur within the recombinagenic complex make the possibility of direct Int/host factor contact seem more relevant.

Figure 49. Model of *attB* Capture

A replica of a band shift lane showing all of the complexes described above is shown on the right. To the immediate left are representative complexes for observed species. Protomers of Int are shown in blue, mIHF in red, and black lines represent DNA. Yellow boxes on DNA represent core and arm-type binding sites. In this model, Int first binds *attP* at P4/P5 as previously described. Int also binds *attP* at P1/P2 forming a bridging complex with *attB* in the presence of *attB* and low concentrations of mIHF, however thi complex is unproductive. The data presented here indicates that as mIHF is added, the P4/P5/Int complex is stabilized as Ints affinity for P4/P5 is improved because of mIHF binding. Further addition of mIHF allows the formation of intasome. mIHF catalyzes the bending of DNA in this complex and also interacts directly with Int to help bring the complex together. This allows the complex to be stable enough to remain intact during EMSA. The result of EMSA with W82L suggests that the addition of mIHF allows even stronger binding of Int to *attP* and *attB* core. This could be due to a conformational change as indicated by the enlarged representation on the far right. If a structural element, shown here as a long oval, is removed from the active site, binding of mIHF to Int might move this element into position. This would both strengthen the Int/core interaction, and activate this protomer of Int. This complex is so similar to intasome that it cannot be resolved as it's own complex in EMSA. Finally, addition of more mIHF allows the binding of P1/P2 to the unbound arm-type binding sites. Binding of Int to P1/P2 allows Int bring *attB*/Int the complex. Binding of mIHF to *attB* at regions flanking the core-binding site could stabilize this interaction. The active complex is formed and recombination can go forward. The P4/P5/*attB* complex may also be stabilized at high mIHF concentrations, but these complexes are non-productive.


20 X

The mutants characterized in this chapter have lost the ability to form specific complexes made by wild type Int. However, they retain the ability to perform recombination. This phenotype varies for each mutant, with A150E having the most severe and W82L having the least severe. This is illustrated by the recombination rates on supercoiled substrates observed for the mutants, and their ability to form intasome at high mIHF concentrations. Recombination efficiency of these mutants is increased with the addition of mIHF. This shows that the stabilization of intasome observed in EMSA translates into increased efficiency *in vitro*. The efficiency of recombination is also improved in the presence of high concentrations of *attB* indicating that *attB* capture may also be affected by these mutations. However, no complexes are formed in the presence of high concentrations of *attB* in the absence of mIHF, and *attB* appears to destabilize intasome formed by these mutants at high mIHF concentrations. Assuming that these mutants bind arm-type binding sites with greater affinity than core as is true for other tyrosine recombinases (Kim and Landy 1992, Minter et al. 1986), the ability of these mutants to form synaptic complex 1 at high mIHF concentrations but not at high *attB* concentrations suggests that mIHF is acting to stabilize Int core interactions on both *attP* and *attB*. If these mutants are unable to bring mIHF into the complex, because they are defective for Int/mIHF, protein/protein interactions, high concentrations of mIHF will increase the incidence of mIHF joining the complex by chance, and the bending of *attP* required to form intasome will only be observed at high mIHF concentrations. However, *attB*/Int interactions also appear to be stabilized at high mIHF concentrations. EMSA indicates that mIHF is able to stabilize the P1/P2/*attB* intermolecular bridge created by W82L since synaptic complex 1 is formed at high mIHF concentration, but not at high *attB* concentrations. If the W82L simply has a lower affinity for *attB*, high concentrations of *attB* should be able to compensate for this

defect, but synaptic complex 1 is not observed at high *attB* concentrations, indicating that there is another deficiency that is causing the mutant phenotype.

W82L, T92A and H138L are able to form a new complex at high mIHF and *attB* concentrations. This complex runs just above intasome at about the same position as the P4/P5/*attB* complex observed in wild type reactions carried out at high concentrations of *attB* in the absence of mIHF. This complex could indicate that the P4/P5/*attB* intermolecular bridging complex is created more frequently at high concentrations of mIHF. In this case, mIHF is stabilizing the Int/core interaction between Int and *attB*. This could mean that these mutants have decreased ability to form the P1/P2 *attB* intermolecular bridge, but it could also indicate that mIHF is not catalyzing the bending of *attP* required to bring *attP* core into position for Int binding. Int is binding P4/P5 that activates the catalytic domain for core binding, but since *attP* core cannot be bound *attB* is bound instead.

A direct interaction between mIHF and Int could increase the affinity of Int for core in two ways. Protein/protein interactions could induce a conformational change that allows Int to bind more tightly to core. Structural data suggests that the tyrosine containing helix of Cre (helix M) moves into position only after the Cre/*LoxP* complex is formed (Van Duyne 2001). If this is true for L5 Int, the helix in this position in L5 Int may move into position only after mIHF has contacted Int. mIHF might also mediate interactions between Int and core by binding directly to both. The inability of L5 Int to utilize other DNA binding proteins that have been implicated in recombination such as IHF or HU in place of mIHF *in vitro* (Lee and Hatfull 1993) suggests that mIHF makes specific contacts to either Int or *attP*. Since mIHF has been shown to bind non-specifically to DNA, it is probably binding directly to Int. Footprinting data clearly shows that mIHF binds to the DNA flanking core only when Int has bound core

suggesting that Int and mIHF bind to core cooperatively. A protein/protein interaction between Int and mIHF would allow mIHF to bend the DNA in the proper place, while utilizing mIHF's intrinsic affinity for DNA to hold the substrate in position. In a third scenario, a protein/protein interaction between Int and mIHF would perform both functions.

Heterobivalent binding between arm-type and core binding sites is essential for formation of recombinationally active complexes. Binding of arm-type sites has been thought to stimulate the binding of core through a series of conformational changes, and the middle domain was thought to mediate this cross talk between the outer domains. The middle domain has also been implicated in core binding. The middle domain mutants characterized here suggest that both of these actions can occur based not on arm-type binding but on binding of mIHF to Int. This leads to a model in which core binds to arm-type binding sites on *attP*. mIHF is recruited to the complex by protein/protein interactions with Int and binds to both *attP* and Int. This interaction allows mIHF to bind to *attP* and catalyze the required DNA bending. mIHF also forms a protein bridge between Int and the *attP* core that holds core in position, and intasome is formed. In the presence of *attB*, mIHF's bridging is also necessary for *attB* binding allowing the synaptic complexes to be formed and stabilized.

V. LYSOGENY OF WILD PHAGE

V.A. Introduction

Mycobacteriophages make up one of the largest groups of isolated phages. The advent of high capacity DNA sequencing instruments and protein and DNA databases has made it possible to do genomic studies on large groups of bacteriophages. Comparisons of protein and DNA sequences within phage genomes give clues to the phages interactions with other phages' and the host. The functions of various genes within the phage genome can be determined by comparing sequences of putative proteins, and the enzymatic tools that the phage uses to take over the host cell during lytic growth, or remain a silent part of the host during lysogeny, can be revealed.

The ability of a bacteriophage to form stable lysogens is the defining feature of many phages. Mycobacteriophage L5 is a good example of a lysogenic bacteriophage that forms an integrated prophage using a phage-encoded integrase that catalyzes recombination between the host chromosome and the phage genome, placing the L5 lysogen within the host chromosome where it can be replicated with the host DNA and passed to daughter cells (Lee and Hatfull 1993). Many lysogenic phages perform similar recombination reactions within their host's genome (Ackermann 1998, Barer and Harwood 1999, Gentry-Weeks et al. 2002, Miao and Miller 1999). Other phages form extrachromosomal prophage that act like low copy number plasmids in the host cell (Mann and Slauch 1997, Park et al. 1998, Xiong et al. 1996). The genomes of these phages usually contain partitioning proteins whose function is to ensure that one plasmid prophage is passed to each daughter cell (Bouet et al. 2000, Erdmann et al. 1999, Grigoriev and Lobocka 2001).

Relative to the number of phage in the biosphere, few bacteriophages have been studied in depth, and fewer lysogeny systems have been isolated and characterized. To study the breadth of lysogeny systems within mycobacteriophages, wild mycobacteriophages were isolated, characterized for lysogeny, and their genomes sequenced. The sequences were then scrutinized using BLAST searches of the protein databases, and putative lysogeny proteins were identified. The method of lysogeny can then be discerned, and the necessary proteins can be isolated and studied.

V.B. Specific Aims

1. Isolate bacteriophage from natural sources that are able to infect *Mycobacterium smegmatis*.
2. Characterize the isolated phages' plaque formation and ability to form lysogens.
3. Sequence isolated phages' genomes, and compare genomic sequences with protein database, to identify putative lysogeny proteins.

V.C. Isolation of Mycobacteriophage CJW1

Mycobacteriophage CJW1 was isolated from the compost pile of Carl and Joann Wadsworth of Edinburgh, Ohio. Bacteriophages were isolated from soil samples taken from the pile by mixing a small amount of the moist compost with ~ 20 mL of phage buffer and thoroughly breaking up the soil by vortexing. Particulate matter was eliminated from the mixture by centrifugation. The supernatant was extracted and filtered through a 0.45 μm syringe filter to eliminate bacterial cells as well as other contaminants are unable to pass through the 0.45 μm filter. One – 100 μL of the phage containting filtrate was then mixed with 1 mL of an overnight *M. smegmatis* culture

grown to near saturation, and the phage were allowed to adsorb for 20 minutes. These cultures were then mixed with 0.5% Middlebrook top agar and plated on 7H10 plates.

Phages that are able to infect *M. smegmatis* were identified by the presence of plaques, clearings on the lawn of bacteria. Phage within the plaques were picked using a 200 μL pipet tip and placed into 100 μL of phage buffer. Phages were plaque purified by plating 5 μL of the phage-containing buffer with 500 μL of an overnight culture of *M. semgmatis* and picking isolated plaques from these plates. This was accomplished at least three times to ensure that only a single phage clone is present in each solution.

Plaque purified phage were then used to infect an overnight culture of *M. smegmatis* to near clearing. Phages were washed from the plate by flooding the plate with 5 mL of phage buffer and incubating the flooded plate at 4°C for several hours. The buffer was collected and filtered through a 0.45 μm syringe filter. The filtered phage lysate was then used to infect 20 large plates of *M. smegmatis* to near clearing. These plates were flooded with 10 mL phage buffer, and phages were dissolved into the buffer for several hours at 4°C. The lysate was collected, and large particles were removed by centrifugation. Phages were precipitated overnight in 10% polyethylene glycol 8000 and 1 M sodium chloride. Precipitated phage were pelleted by centrifugation and resuspended in 10 mL phage buffer that were mixed with cesium chloride to a density of 1.5 g/mL. Phages were then banded by centrifugation in an ultracentrifuge at 126,000 xg for 24 hours at 10°C. The phage band was isolated and high titer phage stocks were dialyzed against phage buffer overnight. The phages were then collected and centrifuged to remove insoluble particles.

V.D. CJW1 has an Icosahedral Head and a Long Tail

Electron microscopy was used to compare the structure of CJW1 phage particles with other mycobacteriophages. High titer phage stocks were incubated on carbon coated copper mesh grids for 2-5 minutes before excess buffer was removed. Grids containing CJW1 phage were then negatively stained with 0.5% Uranyl Acetate for 1 minute, excess stain was removed, and phage were visualized using a Zeiss 902 electron microscope (Figure 50).

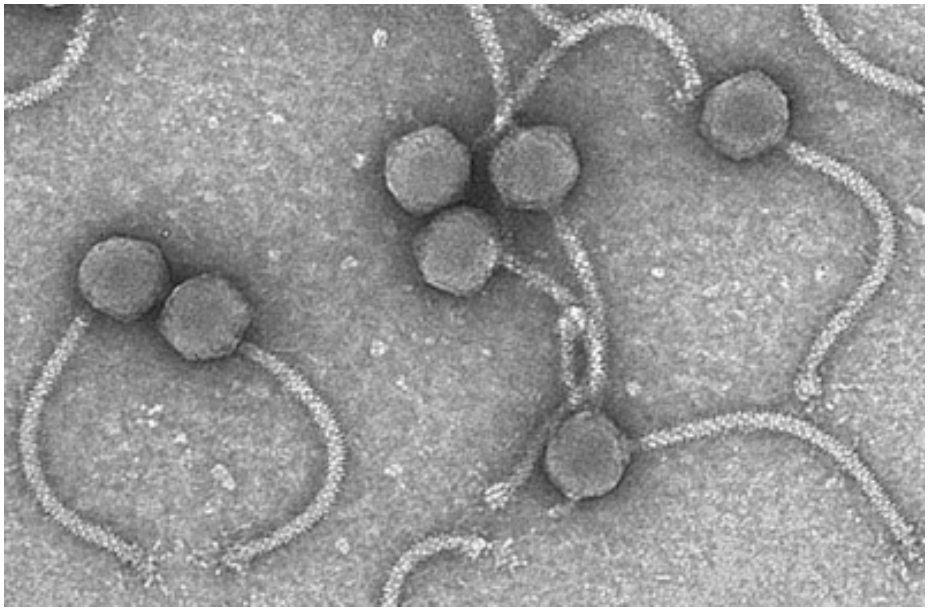
Electron microscopy shows that CJW1 has an icosahedral head that is slightly larger then the head of L5, ~80 versus ~75nM in diameter. The tail of CJW1 is twice as long as L5's, 300 versus 150nM. The tail is flexible with an L5-like foot with finger-like appendages that may be short tail fibers attached at the base of the tail.

V.E. The Genomic Sequence of CJW1

The genomic sequence of CJW1 was fully described by sequencing the entire phage genome. A random shotgun cloning technique was employed to create random sequencable fragments of CJW1. CJW1 DNA was isolated and purified from high titer stocks of CJW1 phage by washing the phage in buffer saturated phenol in a 1:5 ratio of phage to phenol. This effectively destroys the protein coat of the phage while releasing the phage DNA into the buffer. The phenol and protein containing organic phase was then removed, and after several washes, very little protein remains in the solution. Residual phenol was removed by mixing the DNA containing buffer with chloroform. The phenol is dissolved in the chloroform, and can be removed with the organic phase. The buffer is washed with chloroform several times until most of the phenol has been removed. Phage DNA was precipitated by adding 1/10 volume of 3 M sodium acetate

Figure 50. Electron Microscopy of CJW1

Two drops of a 1/10 dilution of high titer phage stocks were placed on a carbon coated copper grid for 2-5 minutes before the remaining liquid was removed. Grids containing CJW1 phage were negatively stained with 0.5% Uranyl Acetate for 1 minute and excess stain was removed. The grids were carefully dried, and the phages attached to the grids were visualized using a Zeus 902 electron microscope.



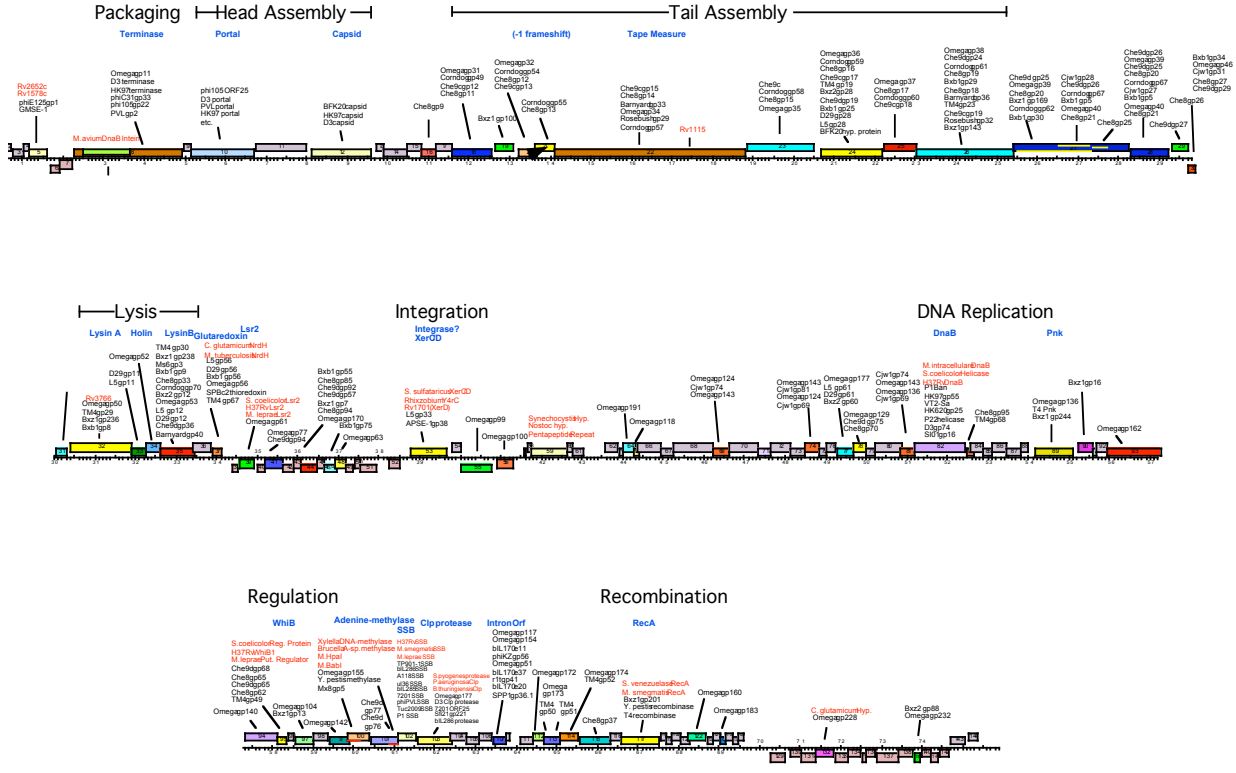
and 3 volumes of 95% ethanol to the remaining aqueous phase of the solution and pelleted by microcentrifugation for 20 minutes at full speed. The DNA pellet was then washed with 70% ethanol and allowed to dry before being resuspended in 50 μL of Tris-EDTA.

Purified DNA was mechanically sheared into small fragments in a GeneMachines Hydroshear, and the ends of the sheared fragments were repaired with Klenow and T4 DNA polymerases with 500 μM dNTPs. The fragments were then run on a 0.7% agarose gel and the 1000 – 3000 base pair fragments were isolated from the gel and cloned into the *EcoRV* site of pBluescript SK^- (Stratagene). Insert-containing plasmids were selected for on X-gal containing LB plates with 12.5 mM tetracycline and 100 mM amplicilin. White colonies were isolated. *E. coli* containing these plasmids were grown in liquid culture overnight, and plasmid minipreps (Qiagen) were performed to purify these plasmids. Purified plasmids were sequenced using the Applied Biosystems Incorporated 3100 Genetic Analyzer from both ends using Big Dye Version 3.0 terminator chemistry to fluorescently label DNA fragments. 1,152 cloned fragments were sequenced in this way, and 70 primers were used to sequence the gaps between contiguous sequences from genomic DNA. Sequences were assembled and edited using the Phred/Phrap/Conced software packages.

Sequencing of the CJW1 genome provided a 75,931 base pair sequence for the CJW1 genome that is 63.1% G+C. Putative open reading frames were derived from this sequence using GeneMark Version 2.4 (available at http://opal.biology.gatech.edu/GeneMark/genemark24.cgi) and GlimmerM (available at http://www.tigr.org/software). The open reading frames were described, and a genetic map was created using DNA Master (J.G. Lawrence)(Figure 51).

Figure 51. Genetic Map of the CJW1 Genome

The genome of CJW1 was sequenced using shotgun cloning of purified CJW1 DNA. Open reading frames were derived using GlimmerM and GeneMark version 2.4 software. Putative open reading frames were cataloged, and the genetic map was created using DNA Master. Genes are shown as multicolored blocks above a line showing the number of base pairs. Genes are staggered to allow easier identification. Each of the putative open reading frames were used as query sequences comparisons of Genbank and PBI protein databases using the BLAST-X algorithm, and gene homologs are shown above the gene in the genetic map.


Packaging
Head Assembly
Tail Assembly
Terminase
Portal
Capsid
(-1 frameshift)
Tape Measure
Lysis
Integration
DNA Replication
Lysin A
Holin
LysinB
Glutaredoxin
Lsr2
Integrase?
XerQD
DnaB
Pnk
Regulation
Recombination
WhiB
Adenine-methylase
SSB
Clp protease
IntronOrf
RecA

Analysis of the CJW1 genomic sequences showed that CJW1 contains 141 tightly packed putative open reading frames with an average size of 546 base pairs. Each of the 141 open reading frames were used as query sequence in BLAST-X searches of the Genbank and Pittsburgh Bacteriophage Institute (PBI) databases using BlastInspector (Xu/Lewis). BLAST-X translates the query DNA sequence into amino acid sequence and uses the amino acid sequence as the query sequence. 55% of CJW1 genes are made up of novel sequences that do not match anything in either the Genbank or PBI databases. 44% match genes that are found in other phage, and of these, only 3.5% are from non-mycobacteriophage.

The CJW1 genome is transcribed almost entirely from a single strand as indicated in Figure 52 with three relatively small and two larger segments of the genome read in the opposite direction. Gene 30 is an individual gene read from the opposite strand, and genes 6 and 7, and 55 and 56 are groups of two genes read in the opposite direction. The clusters of genes between 38 and 52, and 129 and 142 are also read form the opposite strand. These genes could represent morons.

BLAST searches reveal that genes that resemble phage structural genes are clustered on the left side of the genome, and several modules can be discerned based on the arrangment of these genes. The packaging module contains the putative phage terminase, gene 8, followed by the head assembly module with the putative phage portal protein, gene 10, and capsid, gene 12. The tail assembly module contains a putative tape measure protein, gene 22, with a –1 base pair frame shift just upstream between genes 20 and 21, like other phage-encoded tape measure proteins (Hendrix 1988). There also appears to be a lysis module that contains 2 lysin genes flanking a holin, genes 33, 35, and 34, respectively. The remainder of the genome is composed of

genes with some homology to known genes surrounded by unknowns, so no real modules can be picked out.

CJW1 contains several interesting and unexpected genes. Gene 103 appears to be a Clp protease. Clp proteases are multi-component ATP-dependent proteases that account for 80% of the protein degradation that takes place in bacterial cells (Goldberg et al. 1994, Laskowska et al. 1996). Clp proteases have been shown to be especially active against small regulatory proteins such as repressors and enhancers (Mhammedi-Alaoui et al. 1994, Schweder et al. 1996). CJW1 may use this protease to destroy its own repressor, or those of the host to deregulate gene expression during lytic growth.

CJW1 encodes a polynucleotide kinase (PNK) (Folk 1981, Weinfeld et al. 1989) encoded by gene 89 and a DnaB homologue encoded by gene 82. DnaB proteins are heximeric helicases that are utilized in DNA replication (Bhattacharyya and Griep 2000, Johnson et al. 2000, Yu et al. 1996). The presence of a DnaB type helicase indicates that phage encoded enzymes may be necessary for replication of the phage genome. The products of both the DnaB and PNK genes would be active during replication. The proximity of these genes to each other could indicate that there is a cluster of replication genes in this region of the genome.

The CJW1 terminase appears to encode an intein. An intein promotes the posttranslational processing of the protein in which it is housed by catalyzing excision of its own amino acids and the ligation of the flanking protein regions. The splicing reaction is carried out in the absence of cofactors or auxillary enzymes (Xu and Evans 2001). Very few phages have been shown to contain inteins, making CJW1 unusual. The polypeptide sequence of the intein is also unique. Splicing of inteins is thought to require a Cysteine, Serine, or Threonine at the beginning and end of the intein polypeptide sequence. The CJW1 intein closely resembles an intein found in the DnaB

of *Mycobacterium avium* that has an Alanine as the N-terminal amino acid. Alanine substitutions at this position in other studied inteins block the splicing reaction (Evans et al. 1999, Perler 1998). However, the *M. avium* intein appears to be able to catalyze the posttranslational proteolytic processing of the DnaB protein (Yamamoto et al. 2001).

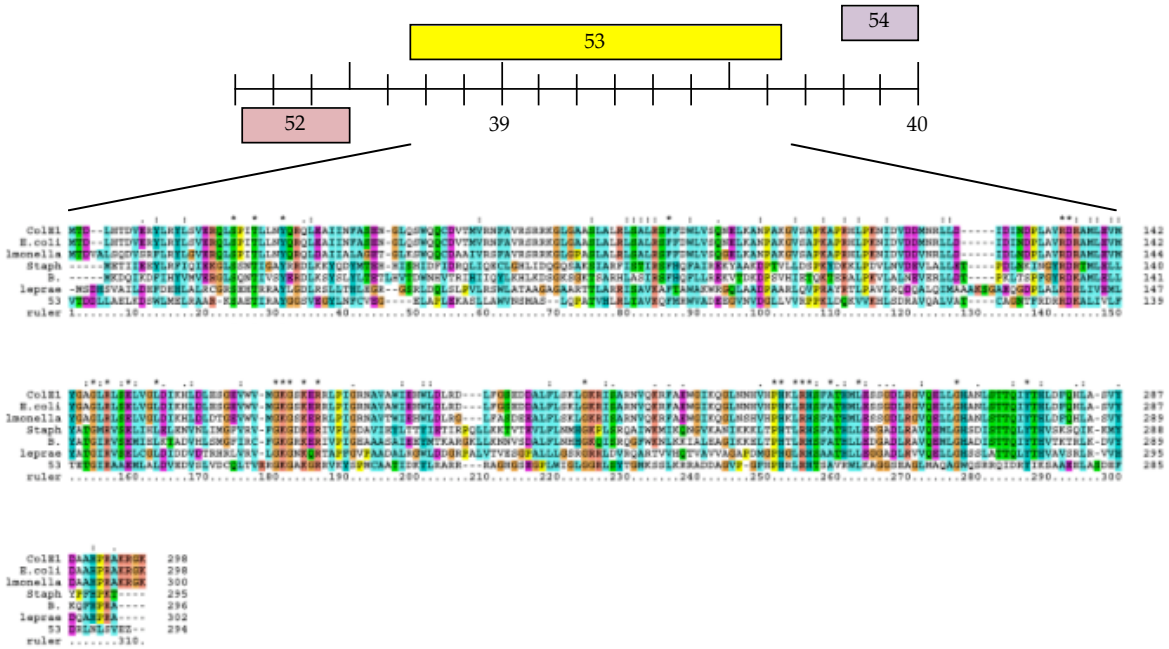
V.F. CJW1 Contains a XerCD-Like Recombinase

Lysogeny for many phages involves integrating the phage genome into the host chromosome. Phage integrases are usually tyrosine recombinases and mediate recombination by binding to a site on the phage genome, the phage attachment site, *attP*, and a site with similar sequence on the host chromosome, the bacterial attachment site, *attB*. Gene 53 of CJW1 shares similarity with XerCD type recombinases (Figure 52). Although XerCD recombinases have traditionally not been implicated in integration, recent evidence suggests that host encoded XerCD may mediate the insertion of the genome of filamentous phage CTX into the *Vibrio cholerae* genome (Huber and Waldor 2002). Since gene 53 is located in the middle of the CJW1 genome and is surrounded by non-coding sequence (Figure 53), it is a good candidate for a phage integrase.

Since XerCD recombinases are known to require cofactors for recombination, the CJW1 genome was closely examined for the presence of proteins similar to PepA and ArgR that are known to be necessary cofactors for XerCD recombination at various sites on the *E. coli* genome (Guhathakurta et al. 1996, Hodgman et al. 1998, Summers 1998). These proteins bind to accessory DNA flanking the overlap region in *dif* and *cer*, XerCD DNA substrates and change the regional architecture of the DNA. The change in architecture allows XerD to join the XerC/substrate complex, and recombination is completed (Colloms et al. 1997, Ferreira et al. 2001). If the product of the CJW1 XerCDgene were utilized in the integration of the CJW1 genome into the host

Figure 52. The CJW1 XerC/XerD Recombinase

The region of the CJW1 genome containing gene 53 is shown as a linear representation above the sequence alignment of gp53 with known XerC/D type recombinases. Non-coding regions are shown as thin black lines, and colored boxes indicate coding sequence. Similar amino acids are shown in the same color, and conserved amino acids are marked with asterisks above the amino acid. Vertical lines above amino acids indicate that the amino acids are conserved in almost all of the tested recombinases. Gp53 has similarity to all of the XerC and XerD tyrosine recombinases tested and contains the active site tyrosine as well as the catalytic R-H-R triad indicative of tyrosine recombinases.


54
53
52
39
40
Col81
E.coli
lmonella
Staph
B.
lepras
53
ruler

chromosome, accessory proteins would probably be necessary. No matches to PepA or ArgR were found, however several putative genes in the CJW1 genome would produce proteins with similarity to DNA binding proteins. Any of the large number of unknown proteins could be used in this capacity, or the CJW1 XerCD could use host-encoded accessory proteins.

Small proteins that bind directly to XerCD protomers can also modulate XerCD recombination. FtsK is the best studied of these proteins. FtsK links chromosome segregation and cell division through its N-terminal domain (Ferreira et al. 2001, Sherratt et al. 2001), and FtsK is essential for resolution of chromosome dimers (Barre et al. 2000, Recchia et al. 1999, Recchia and Sherratt 1999, Steiner et al. 1999). There are no matches to FtsK-like proteins in the CJW1 genome, however CJW1 may sequester host-encoded factors for this purpose or an unidentified gene may activate the recombinase activity of the CJW1 XerCD. Therefore, the ability of the product of gene 53 to perform CJW1 integration cannot be ruled out. Overall, very few conclusions can be drawn based on sequence comparison data, but a XerCD type recombinase that acts as a phage integrase is an interesting proposal.

V.G. CJW1 May Repress Lytic Activity Using the Products of Genes 95 or 100

The presence of a putative integrase in the CJW1 genome would indicate that CJW1 is lysogenic. Early CJW1 infections on lawns of *M. smegmatis* resulted in clear plaques, indicative of a lytic phage. If CJW1 is actually lysogenic, the CJW1 genome should contain a repressor protein. Repressor proteins turn off lytic genes, allowing the majority of phage genes to remain silent during infection. In the absence of the phage repressor, the phage undergoes lytic growth and cannot form a stable lysogen even if the recombinase gene is active (Anderson et al. 1981, Brown et al. 1997, Burz et al. 1994,

Dodd et al. 1993, Flashman 1978, Liu and Haggard-Ljungquist 1999, Peña et al. 1998). The CJW1 genome was searched for possible repressor proteins. Gene 95 has homology to WhiB, a regulatory protein found in *Streptomyces coelicolor* that affects sporulation (Hutter and Dick 1999, Molle et al. 2000, Mulder et al. 1999). Homologues of WhiB have been found in the genomes of *M. tuberculosis, M. leprae* and *M. avium*, and the *M. tuberculosis* homologue has begun to be characterized (Mulder et al. 1999). CJW1 gene 95 has homology to all three of these WhiB genes making it a good candidate for the repressor protein.

Gene 100 of the CJW1 genome encodes an adenine methylase similar to those found in several mycobacterial species. Although the use of methylation as a method of gene regulation is not prevalent in prokaryotes, eukaryotes and some viruses are known to use DNA methylation regulate gene expression. Hypermethylation of CpG islands on genes has been shown to inactivate genes in some cancer cell lines, causing tumor formation and in some cases apoptosis (Brault et al. 2002, Esteller et al. 2002a, Esteller et al. 2002b, Fraga and Esteller 2002). Phages are also thought to utilize methylation as a method of remembering which host has been infected based on the methylation patterns left on the phage DNA after lysis (Casadesus and D'Ari 2002). The ability of CJW1 to specifically methylate its own DNA could allow it to down-regulate phage genes based on the inability of host enzymes to interact with methylated DNA. Methylation could also help target phage-encoded enzymes specifically to CJW1 DNA, if phage-encoded enzymes recognized methylated sites that host enzymes cannot. In either case, lysogeny would depend on the actions of gp100.

V.H. CJW1 Forms Lysogens at 42°C but not at 37°C

The presence of a putative integrase and a repressor-like protein in the CJW1 genome indicates that CJW1 may be a lysogenic phage. However, initial characterization of CJW1 plaque morphology indicates that CJW1 is a lytic phage. To examine the life style of CJW1 more closely, the phage was plated on lawns of *M. smegmatis* grown at different temperatures and on different media. Phage/host interactions are crucial to the lysogenic lifestyle. Many phage-encoded repressors are degraded by specific proteases produced by the host. When the concentration of the protease is increased in the host cell, the repressor is degraded, and the phage is induced (Ackers et al. 1982, Lewis et al. 1983, Ptashne et al. 1976). The activity of many genes is regulated based on environmental factors, and since lysogeny is affected by the state of the host cell, these changes could influence CJW1 lysogeny (Banuett and Herskowitz 1987, Banuett et al. 1986, Noble et al. 1993).

High titer CJW1 phage stocks were diluted 1/1000, and 1 μL of this dilution was used to infect 500 μL of an overnight culture of *M. smegmatis*. Phages were allowed to adsorb for 20 minutes, and infected cells were plated on 7H10 plates using a 0.5% MBTA overlay. 7H10 plates were prepared under various conditions to test the ability of CJW1 to infect *M. smegmatis* in the presence of different carbon sources and at different temperatures. Plates containing 100 mM calcium chloride and albumin dextrose complex enrichment, ADC, 100mM calcium chloride and 1% glycerol, and 100mM calcium chloride and 1% glucose were incubated at 42°C, 37°C, 30°C, and room temperature until plaques could be clearly observed. *M. smegmatis* grows slowly at lower temperatures, so plates grown at room temperature and 30°C were incubated for several days.

The carbon source on which *M. smegmatis* cells was grown had little affect on CJW1 infection, and plaques were observed under all conditions. However at 42ºC, the plaques appeared to be cloudy, indicating that CJW1 was forming lysogens at this temperature, but not at 37ºC, 30ºC, or room temperature (Figure 53).

Streaking the cells within the plaques was used to test for the presence of CJW1 lysogens. Plaques on plates grown at 37ºC and 42ºC were streaked and bacterial cells were grown at the same temperatures. Colonies were detected on plates grown at both 37ºC and 42ºC., but the cells streaked from plates at 42ºC formed more colonies (Figure 54). To remove any exogenous phage, single colonies on these plates were streaked and grown at 37ºC or 42ºC. The density of colonies was decreased each time colonies were streaked, but colonies continued to form at 42ºC and not 37ºC.

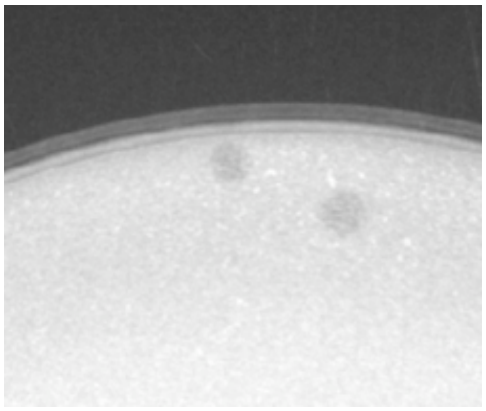
The ability of putative lysogens grown at 42ºC to grow at 37ºC was tested to determine if lysogeny could be maintained at 37ºC after being established at 42ºC. Colonies grown at 42ºC on plates containing putative CJW1 lysogens were streaked onto plates and incubated at 30ºC, 37ºC, and 42ºC (Figure 55). No growth is observed on plates grown at 30ºC or 37ºC indicating that the induction of CJW1 lysogens is absolute at these temperatures. At 42ºC, putative CJW1 lysogens continue to grow normally. Taken together, these results show that CJW1 lysogeny can be neither created nor maintained at 30ºC or 37ºC, and CJW1 lysogens are induced at these temperatures.

After several rounds of streaking and replating, lysogens at 42ºC and 37ºC were tested for the presence of phage. Phage lysogens are induced at some frequency when grown in liquid culture. Therefore, the presence of CJW1 in liquid cultures of lysogens grown at 42ºC would show that CJW1 is forming lysogens at this temperature and not at 37ºC. Liquid cultures were created by picking single colonies from lysogen plates

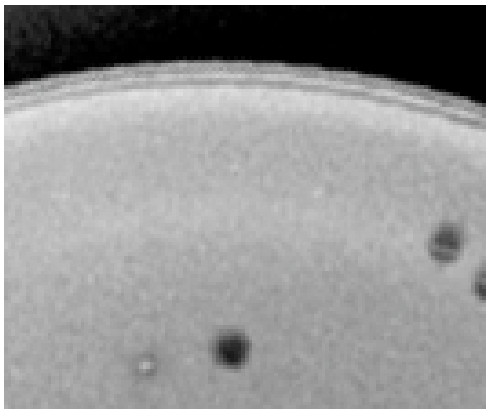
Figure 53. Plaque Morphology of CJW1 Infection at 37°C and 42°C

1/1000 dilution of high titer bacteriophage stocks were mixed with 500 μL of an *M. smegmatis* overnight culture. Phage were allowed to absorb for 20 minutes before being plated in a 5% Middlebrook top agar overlay on plates containing 100 mM calcium chloride and ADC, 100mM calcium chloride and 1% glycerol, and 100mM calcium chloride and 1% glucose were incubated at 42ºC, 37ºC, and 30ºC. Plaque formation was observed 48hrs. The type of media had little effect on plaque morphology, but plates grown at 42ºC showed a cloudier phenotype indicating that lysogeny is occurring.

L5



CJW1-37ºC



CJW1-42ºC

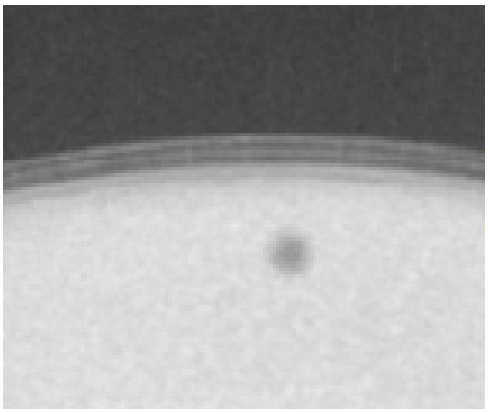


Figure 54. Growth of CJW1 Lysogens at 42°C

Plaques created by CJW1 infection at 42ºC and 37ºC were were streaked onto plates containing 100 mM calcium chloride and ADC, 100mM calcium chloride and 1% glycerol, and 100mM calcium chloride and 1% glucose. These plates were grown at their respective temperatures. Plates grown at 37ºC show little growth of CJW1 lysogens on either media. CJW1 lysogens grow much better at 42ºC as noted by strong growth on streaked area. CJW22 is an isolate of CJW1.

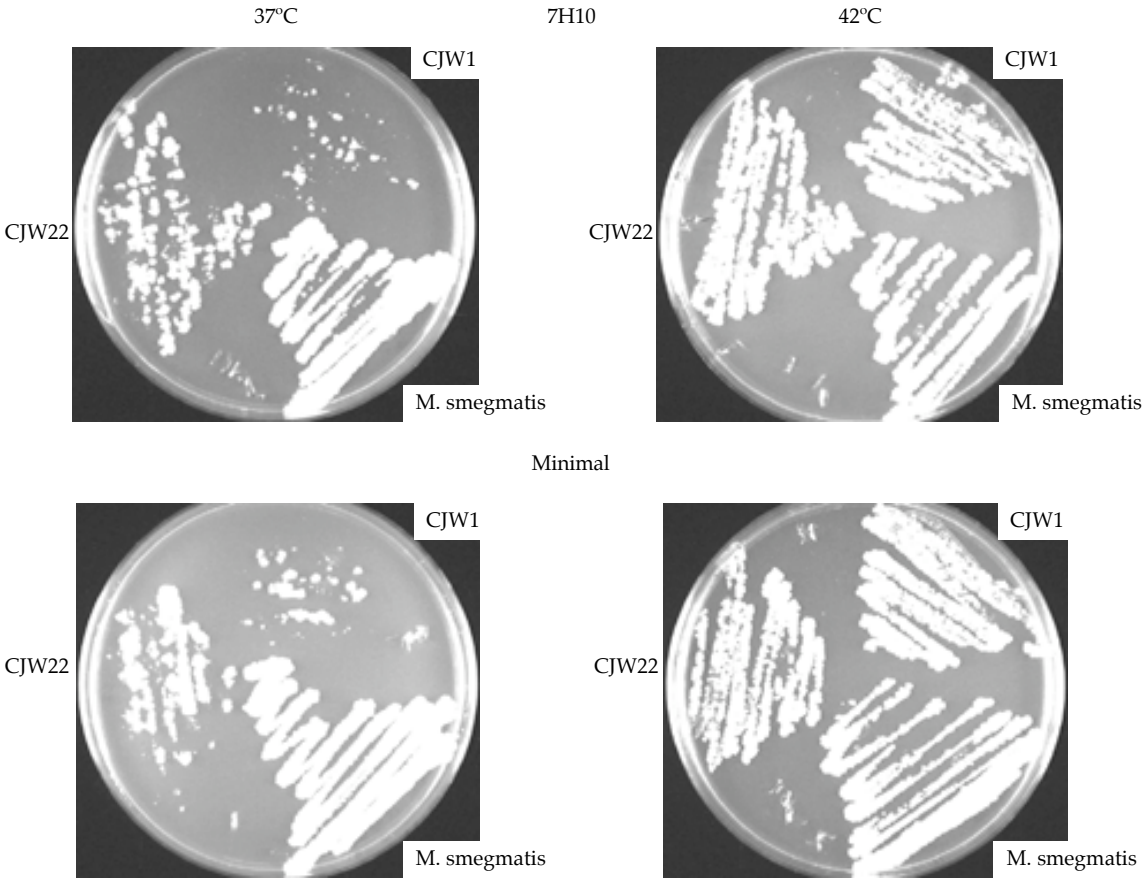
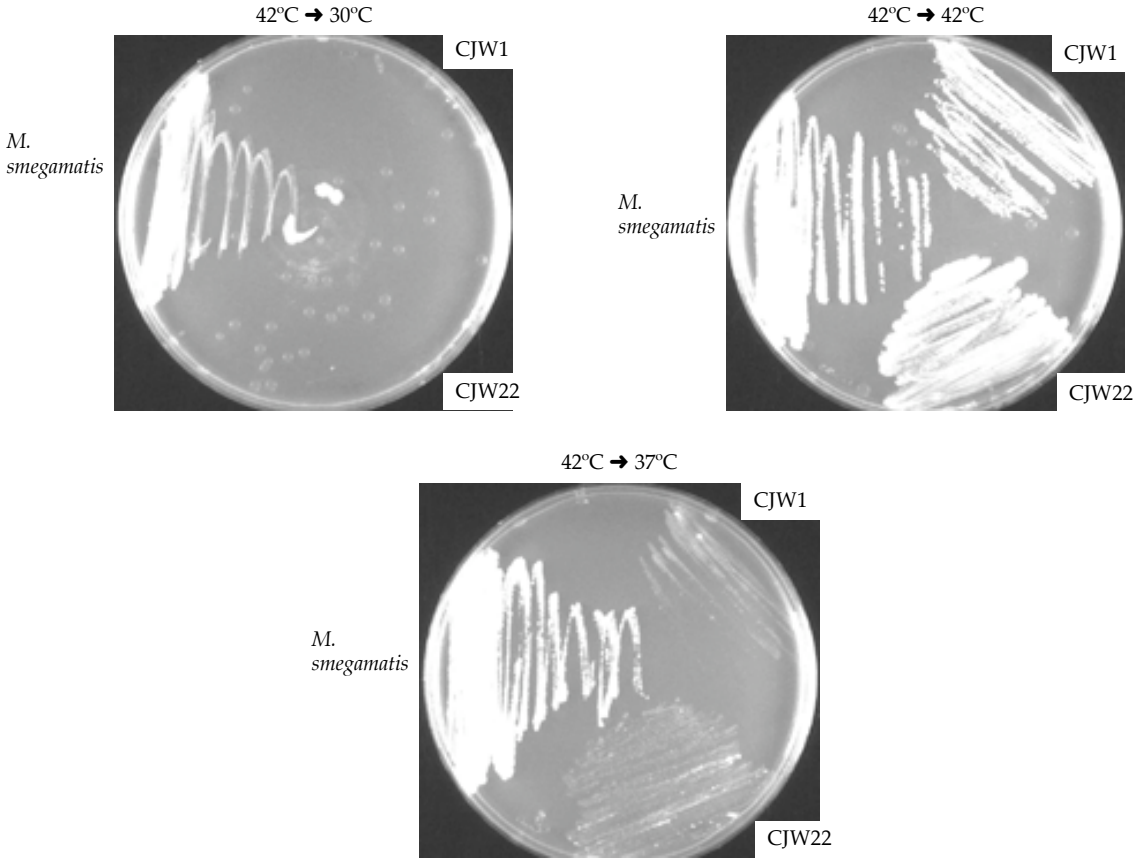

37ºC
7H10
42ºC
CJW1
CJW22
M. smegmatis
CJW1
CJW22
M. smegmatis
Minimal
CJW1
CJW22
M. smegmatis
CJW1
CJW22
M. smegmatis

Figure 55. Maintenance of Lysogeny at 37°C after Establishment at 42°C

Lysogens grown at 42°C and colony purified several times were streaked onto plates containing 100 mM calcium chloride and 1% ADC, and grown at 30°C, 37°C, and 42°C. Plates grown at 30°C and 37°C show very little growth, while plates grown at 42°C continue to grow well.

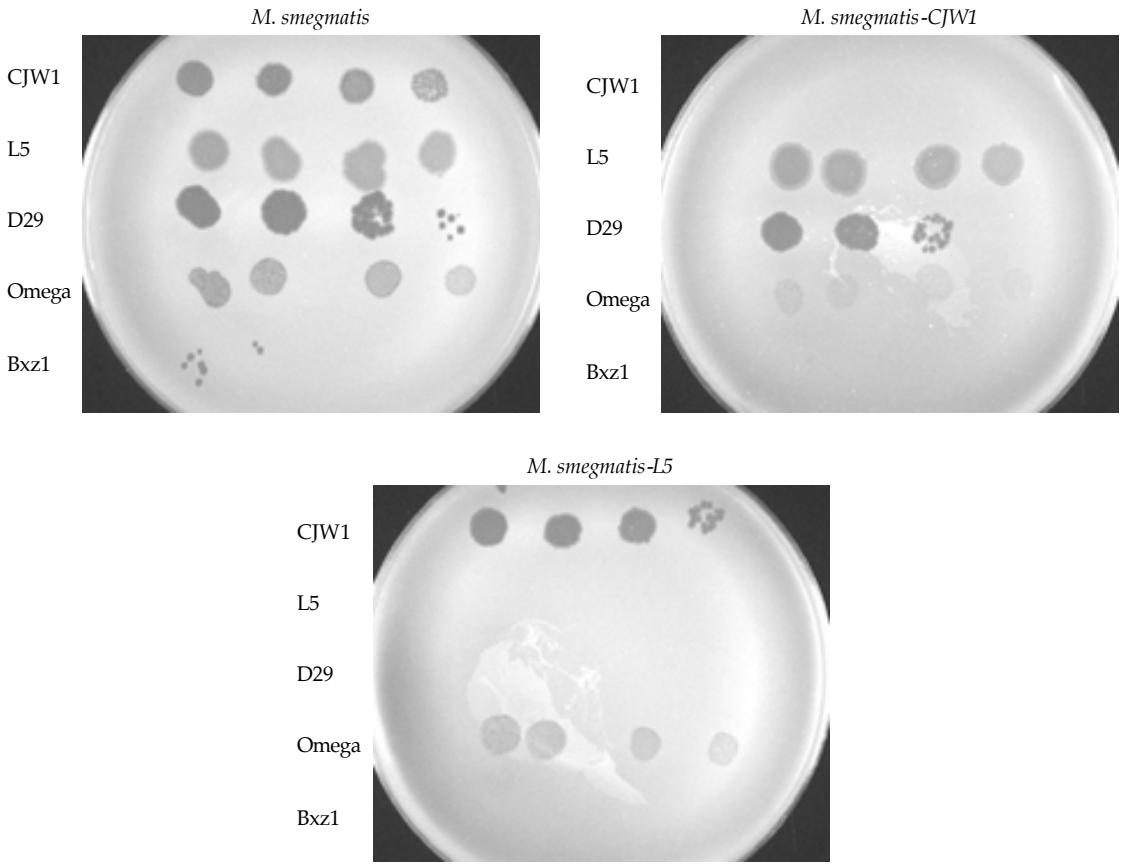

42ºC ➜ 30ºC
CJW1
M. smegamatis
CJW22
42ºC ➜ 42ºC
CJW1
M. smegamatis
CJW22
42ºC ➜ 37ºC
CJW1
M. smegamatis
CJW22

grown at 42ºC and 37ºC into 3 mL of 7H9 media containing ADC and Tween80, and incubating these cultures at 42ºC with shaking for several days. 1 mL of the resulting cultures was then centrifuged to remove bacteria, and the 5 μL of the supernatant was spotted onto plates containing a lawn of *M. smegmatis*. Each of the supernatants from cultures created from putative CJW1 lysogens grown at 42ºC caused plaque formation on *M. smegmatis* lawns, indicating that the colonies grown at 42ºC contained CJW1 prophage. Cultures grown from colonies on plates grown at 37ºC did not contain phage and therefore are not CJW1 lysogens. Phage lysogens are resistant to superinfection by the same phage but susceptible to infection by heterologous phages. Therefore, the presence of CJW1 lysogens can be discerned by challenging putative phage lysogens with CJW1 phage. Putative CJW1 and L5 lysogens as well as non-infected *M. smegmatis* were picked as single colonies from plates grown at 42ºC after several rounds of streaking. These colonies were grown in liquid medium, 7H9 with 1% glycerol and 100mM calcium chloride added, with shaking for 48 hours at 42ºC. 500 μL of these cultures were then plated as lawns on 7H10 plates and serial dilutions of CJW1, L5, D29, Omega, and Bxz1 were spotted onto these plates (Figure 56). Each of the challenge phage produced plaques on the control plate containing uninfected *M. smegmatis*. On L5 lysogen containing plates, plaques formed when CJW1 and Omega were spotted, but no plaques appeared in spots containing L5 or D29, as expected. CJW1 was unable to infect cells on plates containing CJW1 lysogens, however each of the other challenge phage caused plaque formation indicating that the lysogens grown at 42ºC confer superinfection immunity to CJW1 but not other phage. This shows again that CJW1 is able to form stable lysogens at 42ºC.

Colony PCR was used to show that CJW1 lysogens contain a stable CJW1 prophage, and that the lysogens were not immune to CJW1 infection through genetic

Figure 56. Infection of CJW1 on L5 and CJW1 Lysogens

5 μL of serial dilutions (N, 1/10, 1/30, 1/100) of CJW1, L5, D29, Omega and BxzI were spotted onto lawns of *M. smegmatis*, L5 lysogens, and CJW1 lysogens grown at 42ºC on media containing 100 mM calcium chloride and 1%ADC. These plates were incubated for 48 hours until plaques formed. All of the challenge phage infected uninfected *M. smegmatis*. CJW1 is able to infect L5 lysogens, but L5 and D29 are not able to infect these cells. CJW1 is unable to infect CJW1 lysogens indicating that CJW1 lysogeny confers immunity to superinfection. Omega is able to infect all of the tested bacteria, and BxzI is unable to infect either lysogen. The inability of BxzI to infect L5 or CJW1 lysogens is probably due to low phage density in the BxzI stock.


M. smegmatis
CJW1
L5
D29
Omega
Bxz1
M. smegmatis-CJW1
CJW1
L5
D29
Omega
Bxz1
M. smegmatis-L5
CJW1
L5
D29
Omega
Bxz1

mutation in the host cell. 1 μM primers specific to CJW1 were mixed with a single CJW1 lysogen colony under standard PCR conditions, and PFU was used to amplify a 635 base pair fragment within the CJW1 genome. PCR was carried out for 20 rounds, and the products were separated on a 1% agarose gel. Results were visualized by ethidium bromide staining under UV light (Figure 57). A 635 base pair band was observed in both a control lane containing purified CJW1 DNA and 2 lanes containing CJW1 lysogens grown at 42°C, but not lanes containing uninfected *M. smegmatis* colonies. These data clearly show that CJW1 lysogens grown at 42°C contain a stable CJW1 prophage.

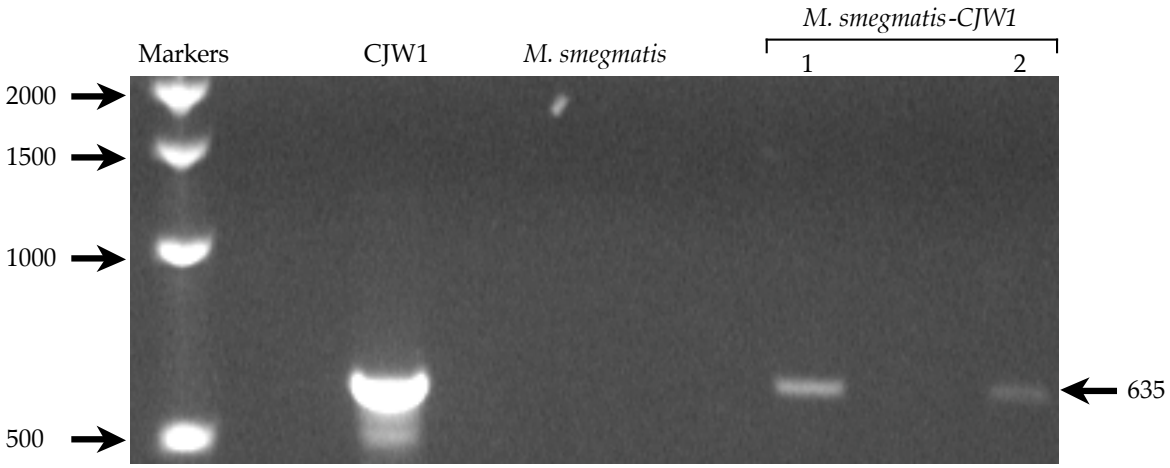
The ability of CJW1 to form lysogens at high temperature is novel among characterized phage. A temperature sensitive repressor that is only active at high temperatures could be the cause this phenotype. Alternatively, the host cell could be producing a protein that blocks the action of the repressor at 30°C and 37°C that is not produced at higher temperatures. The opposite could also be true. Something being produced by the host cell could allow lysogeny in cells grown at 42°C but not at lower temperatures. Temperature sensitive mutations that cause activity at high temperature are rare. CJW1 may form lysogens in host cells that are expressing specific genes, showing the importance of the host/phage interactions.

V.I. Discussion

The "decision" between lysogenic and lytic growth is a complicated process. Genes required for both the lytic and lysogenic pathways in phage □ are expressed upon infection, and the establishment of lysogeny is based on competitive binding between the repressor, cI, and the cro protein, which represses the transcription of cI. Therefore, the level of protein expression of these two proteins is a major determinant in

Figure 57. Colony PCR of CJW1 Lysogens

One μM primers specific to CJW1 were mixed with a single CJW1 lysogen colony under standard PCR conditions, and PFU was used to amplify a 635 base pair fragment within the CJW1 genome. PCR was carried out for 20 rounds, and the products were separated on a 1% agarose gel. Results were visualized by ethidium bromide staining under UV light. A 635 base pair band is observed in both lanes containing CJW1 lysogens as well as control lanes containing purified CJW1 DNA. Reactions containing uninfected *M. smegmatis* did not produce the 635 base pair band.


Markers
CJW1
M. smegmatis
M. smegmatis-CJW1
1
2
2000
1500
1000
500
635

the establishment of lysogeny. cI levels are controlled by a variety of factors including expression of other phage genes, cII and cIII, and the expression and stability of these proteins are controlled by host factors that are effected by the external environment and growth rate of the bacterial host (Ptashne 1987).

Bacterial cells that are grown in rich media are more susceptible to lytic infection by bacteriophage, and lysogenization occurs more readily in cells that have been starved prior to infection (Kourilsky 1973). This is thought to occur through indirect means due to the levels of 3'-5'-cAMP in the host cell (Hong et al. 1971). When cAMP levels are low, as when bacterial cells are grown in rich media, the lytic growth ensues. Under stressful conditions, the cAMP levels are high, and the phage undergoes lysogeny. As a stably integrated prophage, the phage can remain silently attached to the host until the host is growing in more favorable conditions, at which point it can carry out lytic growth, destroying the host cell and infecting other hosts. This allows the phage to carry out lytic growth only under conditions where the host is replicating, so there are new hosts to infect after lysis.

The ratio of phage to bacterial host also affects the decision between lysis and lysogeny. When a large number of phage are present, the lysogenic pathway is favored (Boyd 1951, Lieb 1953). Multiple infections by the same host would increase the cellular level of repressor, pushing the phage toward lysogenic growth. A high phage to host ratio would leave the phage with few hosts to be infected by progeny phage, so establishing lysogeny ensures that progeny will only be produced when the phage to host ratio has been lowered.

The environment in which phage infection is carried out has important implications in the establishment of lysogeny. Phage that are propagated in the lab are grown on hosts that have been grown in rich media under conditions that have been

tested for maximal growth rates. The host is, therefore, growing well, and phages are added at a concentration that allows the formation of single plaques (low phage to host ratio). Mycobacteriophages are commonly isolated from dirt, and/or fecal matter, where conditions and, therefore, the state of the host cell change drastically. In Northeastern Ohio, where CJW1 was isolated, temperatures typically range from –18°C to 40°C during the course of the year. The moisture level also changes drastically from relatively dry in the winter and summer to wet in the spring and fall. The bacterial host of CJW1 must be affected by these changes and compensate by increasing the growth rate when the conditions are good and slowing down the cell,s metabolism or going dormant under poor growth conditions. The lifecycle of a "smart" phage would be tightly coupled to changes in the host metabolism, lysing cells when the host is growing well and undergoing lysogenization when the host is dividing more slowly or is dormant. Continuing to lyse dormant bacteria would eradicate the host population and end the phage lifecycle since the parasitic lifestyle requires viable host for replication.

The ability of CJW1 to lysogenize cells grown at 42°C but not 37°C reflects the tight coupling of the host state to the phage lifestyle. The change in temperature is probably accompanied by changes in the expression pattern of certain genes since the bacteria must compensate for the change in the environment. Changes in metabolism can also cause increases in the level of cofactors like cAMP that are produced as a result of stress response. CJW1 could sense these changes through interactions with cofactors, as is the case in cII, or proteins that are expressed in cells grown at 42°C that are not expressed in cells grown are lower temperatures. Sensing changes involved in host cell stress response would allow CJW1 to form lysogens in cells that are growing more slowly, while undergoing lytic growth in cells that are growing optimally.

VI. CONCLUSIONS

A parasite, by definition, depends on its host for survival. The bacteriophage, as a parasite, requires a bacterial host for replication machinery. However, the phage also employs host proteins to sense the environment of the host cell and decide whether to replicate and destroy the host or lysogenize the cell delaying replication until the growth conditions of the host are more optimal. In the simplest system, phage infection leads directly to the creation of progeny phage and host lysis, lytic growth. A strictly lytic phage, could decimate an isolated population of bacteria relatively quickly leaving the phage without a host. If the host population can be easily replaced, lytic growth can be tolerated. However in isolated ecosystems, lytic growth destroys the population of host that cannot be replaced, so phages have developed a system to sense the state of the host and allow the phage to defer lytic growth until conditions improve, lysogeny.

Mycobacteria are generally considered soil bacteria, and most of the recently isolated mycobacteriophage, like CJW1, have come from soil samples. Non-motile bacteria would be unable to move into or out of most soil ecosystems easily, so the populations of bacteria that the phage infects are limited. Therefore, the phage must be responsive to the growth conditions of the host during infection. Establishing lysogeny while the host is growing well would waste valuable reproduction time, while lysing hosts that are not growing well would quickly destroy the host population. Bacteriophage isolated from environments where the host population can be replaced readily or where phage and bacteria move freely between populations, like laboratory or marine environments, can be less receptive to the state of the host than bacteriophage isolated from soil samples. These phages can afford to undergo lysogeny at high

frequency, since there is a seemingly endless supply of host and are responsive to host cell damage more than growth conditions.

The "decision" between lysogenic and lytic growth is a complicated process. Genes required for lytic and lysogenic growth in phage λ are initially expressed upon infection (Banuett and Herskowitz 1987), and the establishment of lysogeny is based on competitive binding between the repressor, cI (Echols and Guarneros 1983, Johnson et al. 1981), and the cro protein (Georgiou et al. 1979), which represses the transcription of cI. Therefore, the level of protein expression of these two proteins is a major determinant in the establishment of lysogeny. cI levels are controlled by a variety of factors including expression of other phage genes, cII (Shimatake and Rosenberg 1981) and cIII (Echols and Green 1971, Hoyt et al. 1982), and the expression and stability of these proteins are controlled by host factors that are effected by the state of the host cell such as cAMP levels (Banuett et al. 1986, Belfort and Wulff 1973, Unger and Clark 1972).

Integration is also depends on host encoded proteins. λInt requires integration host factor (IHF) to form the recombinagenic nucleoprotein complex with *attP* necessary for integration. In the absence of IHF, other host encoded DNA binding proteins, HU, allow low frequency recombination *in vitro* (Goodman et al. 1992, Segall et al. 1994). Unproductive infection occurs when phage infect bacterial cells that have mutations in the IHF gene, *himA*. The phage attempts to lysogenize the cell, but DNA replication is inhibited as a result of the infection. The host is killed and no phages are release (Mozola et al. 1985). Both the host and the phage depend on functioning host factor during lysogeny.

Chapter IV shows the importance of mIHF in the establishment of L5 lysogeny illustrating that host factors do more then bend DNA. mIHF joins the L5

recombinagenic complex and stabilizes the Int/core interaction allowing the intimate contact between integrase and the recombination substrates. Direct interactions between Int and mIHF may also activate the catalytic domain of Int. These properties appear to be inherent to only mIHF since other similar DNA binding proteins, IHF and HU, cannot be substituted for mIHF *in vitro* (Lee and Hatfull 1993). These results suggest that even if L5 is able to repress lytic gene expression lysogeny cannot occur without functioning mIHF.

CJW1 lysogeny will almost certainly require host factors to create the proper *attP* DNA topology, if CJW1 uses its XerCD type recombinase to perform integration. No proteins within the phage genome appear to contain amino acid similarity to factors that are known to associate with XerCD during recombination, but CJW1 may utilize host-encoded proteins to perform this function. The temperature sensitive phenotype of CJW1 could be explained by the expression of host factors involved in lysogeny at 42°C but not at 37°C. If the phage is able to repress lytic genes through the action of a phage-encoded repressor protein at either temperature, the inability to integrate because of lack of host factor could push the phage toward lytic growth at 37°C. The "decision" between lysogeny and lytic growth would not be based on the phages ability to express more repressor than the host can degrade but on the presence of host proteins that are only expressed at 42°C. CJW1 would therefore be more in tune to the host than lambda or L5 and better suited to survive in harsh environments.

The nature of the parasitic lifestyle requires that the host and the parasite interact with each other intimately. Bacteria and phage depend on each other to maintain a viable ecosystem. The phage population limits the number of bacteria in the ecosystem allowing the bacterial population to be maintained for longer periods of time in environments were resources are a limiting factor. When resources are scarce and

conditions are suboptimal, the bacteriophage undergoes lysogeny, protecting the host from lytic infection by other phage allowing both the phage and host to survive until conditions improve. When conditions improve, lytic growth ensues, and again, phages limit the number of bacteria in the ecosystem. The ecosystem is maintained, and both host and parasite survive for another generation.

APPENDIX

Appendix

A History of Phage Biology

The first description of bacteriophage infection was made in 1915 when F.W. Twort noticed "glassy transformation", plaques, on lawns of bacteria (Twort 1915). F. d'Herelle made similar observations in 1917 coining the term bacteriophage for the corpuscles that caused the plaques (d'Hérelle 1917). D'Herelle continued to study bacteriophage amid controversy that the observations made several years earlier by Twort were caused by similar corpuscles, and in 1926, he published a paper describing the lytic lifestyle, "the corpuscle penetrates the interior of the bacterial cell. When, as a result of its faculty of multiplication, the bacteria corpuscle which penetrated into the bacterium forms a colony of a number of elements, the bacterium ruptures, suddenly, liberating the young corpuscles" (d'Hérelle 1926).

The study of phage continued. Burnett showed in 1929 that 20 –100 phage are released 20 minutes after bacteriophage infection (Burnet 1929). Krueger and Schlesinger independently defined the initial step of phage infection, phage binding to the outside of the bacterial cell, as fixation, or absorption in 1931 and 1932 (Krueger 1931, Schlesinger 1932a, Schlesinger 1932b). The tail was shown to be the organ of absorption and it was shown that tail fibers interact with receptors on the surface of the bacterial cell in 1949 (Anderson 1949).

The modern era of phage biology began in 1939 when Ellis and Delbruck described the first single burst experiment (Ellis and Dellbrück 1939). Delbruck went on

to describe lysis-from-within in 1940 showing that the onset of lysis signals the end of the latent period, the period when the phage reproduces inside the host cell (Dellbrück 1940). Doermann further defined the latent period by breaking open infected cells during the course of infection and looking for infectious phage. In 1952, he described the eclipse period as the time when the infected cell does not contain infectious material (Doermann 1952), and he later described the vegetative phase as the period when the phage multiplies in a non-infectious form during the eclipse period (Doermann 1953).

Schlesinger described the composition of phage in 1934 showing that the phage particle is 50% protein and 50% DNA (Schlesinger 1934), and it was later shown that the DNA resides in the head (Anderson 1950, Herriot 1951). Hershey and Chase performed their well-noted experiments showing that the protein and DNA have independent functions during phage infection, and that DNA is the unit of heredity in 1952 (Hershey and Chase 1952). The production of enzymes encoded on the phage genome were then shown to be necessary for phage replication (Kornberg et al. 1959), and Meselson and Stahl used phage to show semi-conservative DNA replication in 1957 and 1958 (Meselson and Stahl 1958, Meselson et al. 1957).

Recombination was shown to occur between mutant phage in 1946 (Dellbrück and Bailey 1946, Hershey 1946). Hershey and Rotman produced the first genetic map in 1949 using the recombination frequencies mutant genes (Hershey and Rotman 1949), and Benzer used this technique to create fine linkage maps using the frequency of rare recombination events (Benzer 1955) that would lead to his description of point mutations and the cistron as the genetic unit of function in 1957 (Benzer 1957). The ability to map phage genomes eventually led Streisinger, Edgar, and Denhardt to the conclusion that the phage genome is circular (Streisinger et al. 1964). Hershey and Chase showed that 2% of phage particles contain homologous loci from both phage in

the cross (Hershey and Chase 1951), and in 1954, Leventhal offered the partial replica or copy choice hypothesis to explain the recombinant phage (Levinthal 1954). However, Meselson and Weigle described breakage and reunion as the method of recombination seven years later (Meselson and Weigle 1961)

Lysogeny, the ability of some phage to remain latent for several generations, was observed in 1921 by two groups (Bordet and Ciuca 1921, Gildemeister 1921). Lysogenic strains of bacteria were shown to harbor and maintain non-infectious phage, prophage, and it was shown that the prophage could be induced to a vegetative state by external factors in 1951 (Lwoff and Gutmann 1950, Lwoff et al. 1950). Lederberg and Wollman described integrative recombination independently in 1953 (Lederberg and Lederberg 1953, Wollman 1953), and Wollman and Jacob went on to show that integrative recombination occurs site-specifically (Wollman and Jacob 1954, Wollman and Jacob 1957). Jacob also showed that transducing lysogenic phages can carry bits of the host chromosome (Jacob 1955). Superinfection immunity, the ability of lysogenic phage to block infection of other phage, was first described by Kaiser and Jacob (Kaiser and Jacob 1957). They also showed that the C-region of the phage genome was necessary for both immune specificity and lysogeny. Jacob and Campbell then described a "repressor substance" encoded within the C-region in 1959 (Jacob and Campbell 1959).

BIBLIOGRAPHY

Bibliography

Abremski, K., and S. Gottesman. 1982. Purification of the bacteriophage lambda xis gene product required for lambda excisive recombination. *J Biol Chem* 257: 9658-62.

Abremski, K., and R. Hoess. 1984. Bacteriophage P1 site specific recombination. Purification and properties of the Cre recombinase protein. *J Biol Chem* 259: 1509-1514.

Ackermann, H. W. 1998. Tailed bacteriophages: the order caudovirales. *Adv Virus Res* 51: 135-201.

Ackers, G. K., A. D. Johnson, and M. A. Shea. 1982. Quantitative model for gene regulation by lambda phage repressor. *Proc Natl Acad Sci U S A* 79: 1129-33.

Alen, C., S. D.J., and S. D. Colloms. 1997. Direct Interactions of amino peptidase A with recombination site DNA in Xer site specific recombination. *Embo J* 16: 5188-5197.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *J Mol Biol* 215: 403-10.

Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25: 3389-402.

Anderson, T. F. 1949. The reaction of bacterial viruses with their host cells. *Botan. Rev.* 15: 464.

—. 1950. Destruction of bacterial viruses by osmotic shock. *J. Appl. Phys.* 21: 70.

Anderson, W. F., D. H. Ohlendorf, Y. Takeda, and B. W. Matthews. 1981. Structure of the cro repressor from bacteriophage lambda and its interaction with DNA. *Nature* 290: 754-8.

Arciszewska, L. K., R. A. Baker, B. Hallet, and D. J. Sherratt. 2000. Coordinated control of XerC and XerD catalytic activities during Holliday junction resolution. *J Mol Biol* 299: 391-403.

Argos, P., A. Landy, K. Abremski, J. B. Egan, E. Haggard-Ljungquist, R. H. Hoess, M. L. Kahn, B. Kalionis, S. V. Narayana, L. S. d. Pierson, and et al. 1986. The integrase family of site-specific recombinases: regional similarities and global diversity. *Embo J* 5: 433-40.

Banuett, F., and I. Herskowitz. 1987. Identification of polypeptides encoded by an Escherichia coli locus (hflA) that governs the lysis-lysogeny decision of bacteriophage lambda. *J Bacteriol* 169: 4076-85.

Banuett, F., M. A. Hoyt, L. McFarlane, H. Echols, and I. Herskowitz. 1986. hflB, a new Escherichia coli locus regulating lysogeny and the level of bacteriophage lambda cII protein. *J Mol Biol* 187: 213-24.

Barer, M. R., and C. R. Harwood. 1999. Bacterial viability and culturability. *Adv Microb Physiol* 41: 93-137.

Barre, F. X., M. Aroyo, S. D. Colloms, A. Helfrich, F. Cornet, and D. J. Sherratt. 2000. FtsK functions in the processing of a Holliday junction intermediate during bacterial chromosome segregation. *Genes Dev* 14: 2976-88.
Belfort, M., and D. L. Wulff. 1973. An analysis of the process of infection and induction of E. coli hfl-1 by bacteriophage lambda. *Virology* 55.
Benzer, S. 1955. Fine structure of a genetic region in bacteriophage. *Proc Natl Acad Sci U S A* 41: 344.
—. 1957. The elementary units of heredity in W. D. McElroy and B. Glass, eds. *The Chemical Basis of Heredity*. The Johns Hopkins Press, Baltimore.
Better, M., C. Lu, R. C. Williams, and H. Echols. 1982. Site-specific DNA condensation and pairing mediated by the Int protein of bacteriophage lambda. *Proc Natl Acad Sci U S A* 79: 5837-5841.
Better, M., S. Wickner, J. Auerbach, and H. Echols. 1983. Role of the Xis protein of bacteriophage lambda in a specific reactive complex at the attR prophage attachment site. *Cell* 32: 161-8.
Bhattacharyya, S., and M. A. Griep. 2000. DnaB helicase affects the initiation specificity of Escherichia coli primase on single-stranded DNA templates. *Biochemistry* 39: 745-52.
Bode, V. C., and A. D. Kaiser. 1965. Changes in the structure and activity of lambda DNA in a superinfected immune bacterium. *J. Mol. Biol.* 14: 399.
Bordet, J., and M. Ciuca. 1921. Déterminisme leucocytaires et autolyse microbienne transmissible. *Compt. rend. Soc. Sci* 83: 1296.
Bouet, J. Y., J. A. Surtees, and B. E. Funnell. 2000. Stoichiometry of P1 plasmid partition complexes. *J Biol Chem* 275: 8213-9.
Boyd, J. 1951. "Excessive" dose phenomenon in virus infection. *Nature* 167: 1061.
Brault, V., S. Pfeffer, M. Erdinger, J. Mutterer, and V. Ziegler-Graff. 2002. Virus-induced gene silencing in transgenic plants expressing the minor capsid protein of Beet western yellows virus. *Mol Plant Microbe Interact* 15: 799-807.
Brown, K. L., G. J. Sarkis, C. Wadsworth, and G. F. Hatfull. 1997. Transcriptional silencing by the mycobacteriophage L5 repressor. *Embo J* 16: 5914-21.
Burgi, E., A. D. Hershey, and L. Ingraham. 1966. Preferred breakage points in T5 DNA molecules subjected to shear. *Virology* 28: 11-14.
Burnet, F. M. 1929. A method for the study of bacteriophage multiplication in broth. *Brit. J. Expt. Path.* 10: 109.
Burz, D. S., D. Beckett, N. Benson, and G. K. Ackers. 1994. Self-assembly of bacteriophage lambda cI repressor: effects of single- site mutations on the monomer-dimer equilibrium. *Biochemistry* 33: 8399-405.
Cao, Y., and F. Hayes. 1999. A newly identified, essential catalytic residue in a critical secondary structure element in the integrase family of site specific recombinases is conserved in a similar element in eukaryotic type 1B topoisomerase. *J Mol Biol* 289: 517-527.
Casadesus, J., and R. D'Ari. 2002. Memory in bacteria and phage. *Bioessays* 24: 512-8.
Chen, Y., U. Narendra, E. L. Iype, M. M. Cox, and P. A. Rice. 2000. Crystal structure of Flp recombinase-Holliday junction complex: assembly of an active oligomer by helix swapping. *Mol Cell* 6: 885-897.
Cheng, C., P. Kussie, N. Pavletich, and S. Shuman. 1998. Conservative structure and mechanism between eukaryotic topoisomerase I and site specific recombinases. *Cell* 92: 841-850.

Cho, E. H., R. Alcaraz, Jr., R. I. Gumport, and J. F. Gardner. 2000. Characterization of bacteriophage lambda excisionase mutants defective in DNA binding. *J Bacteriol* 182: 5807-12.
Cho, E. H., R. I. Gumport, and J. F. Gardner. 2002. Interactions between integrase and excisionase in the phage lambda excisive nucleoprotein complex. *J Bact* 184: 5200-5203.
Clark-Curtiss, J. E., J. E. Thole, M. Sathish, B. A. Bosecker, S. Sela, E. F. de Carvalho, and R. E. Esser. 1990. Protein antigens of Mycobacterium leprae. *Res Microbiol* 141: 859-71.
Colloms, S. D., J. Bath, and S. D.J. 1997. Topological selectivity in Xer site-specific recombination. *Cell* 88: 855-864.
Craig, N. L. 2002. Mobile DNA: An Introduction in C. N.L., C. R, G. M, and L. A.M., eds. *Mobile DNA II*. ASM Press, Washington, D.C.
Craig, N. L., and J. W. Roberts. 1980. E. coli recA protein-directed cleavage of phage lambda repressor requires polynucleotide. *Nature* 283: 26-30.
d'Hérelle, F. 1917. Sur un microbe invisible atagoniste des bacilles dysentériques. *Compt. rend. Acad. Sci.* 165: 373.
—. 1926. *The Bacteriophage and its Behaviour*. Williams and Wilkins, Baltimore.
Deho, G., D. Ghisotti, P. Alano, S. Zangrossi, M. G. Borrello, and G. Sironi. 1984. Plasmid mode of propagation of the genetic element P4. *J Mol Biol* 178: 191-207.
Dellbrück, M. 1940. The growth of bacteriophage and lysis of the host. *J. Gen. Physiol.* 23: 643-660.
Dellbrück, M., and W. T. Bailey. 1946. Induced mutations in bacterial viruses. *Cold Spring Harbor Symp. Quant. Biol.* 11: 33.
Dodd, I. B., M. R. Reed, and J. B. Egan. 1993. The Cro-like Apl repressor of coliphage 186 is required for prophage excision and binds near the phage attachment site. *Mol Microbiol* 10: 1139-50.
Doermann, A. H. 1952. The intracellualr growth of bacteriophaegs I. Liberation of intracellular bacteriophage T4 by premature lysis with another phage or with cyanide. *J. Gen. Physiol.* 35: 645-656.
—. 1953. The vegetative state in the life cycle of the bacteriophage: Evidence for its occurrence and its genetic characterization. *Cold Spring Harbor Symp. Quant. Biol.* 18: 3.
Doke, S. 1960. Studies on Mycobacteriophages and Lysogenic Mycobacteria. *J Kumamoto Med Soc* 34: 1360-1373.
Dorgai, L., S. Sloan, and R. A. Weisberg. 1998. Recognition of core binding sites by bacteriophage integrases. *J Mol Biol* 277: 1059-70.
Duckett, D. R., A. I. Murchie, S. Diekmann, E. von Kitzing, B. Kemper, and D. M. Lilley. 1988. The structure of the Holliday junction, and its resolution. *Cell* 55: 79-89.
Echols, H., and L. Green. 1971. Establishment and mainteneance of repression by bacteriophage lambda: The role of cI, cII, and cIII proteins. *Proc Natl Acad Sci U S A* 68: 2190.
Echols, H., and G. Guarneros. 1983. Control of integration and excision. Pages 75-92 in R. W. Hendrix, J. W. Roberts, F. W. Stahl, and R. A. Weisberg, eds. *Lambda II*. Cold Spring Harbor Publications, Cold Spring Harbor, New York.
Ellis, E. L., and M. Dellbrück. 1939. The growth of bacteriophage. *J. Gen. Physiol.* 22.
Endemann, H., P. Bross, and I. Rasched. 1992. The adsorption protein of phage IKe. Localization by deletion mutagenesis of domains involved in infectivity. *Mol Microbiol* 6: 471-8.

Endemann, H., V. Gailus, and I. Rasched. 1993. Interchangeability of the adsorption proteins of bacteriophages Ff and IKe. *J Virol* 67: 3332-7.
Erdmann, N., T. Petroff, and B. E. Funnell. 1999. Intracellular localization of P1 ParB protein depends on ParA and parS. *Proc Natl Acad Sci U S A* 96: 14905-10.
Esteller, M., M. F. Fraga, M. F. Paz, E. Campo, D. Colomer, F. J. Novo, M. J. Calasanz, O. Galm, M. Guo, J. Benitez, and J. G. Herman. 2002a. Cancer epigenetics and methylation. *Science* 297: 1807-8; discussion 1807-8.
Esteller, M., M. Guo, V. Moreno, M. A. Peinado, G. Capella, O. Galm, S. B. Baylin, and J. G. Herman. 2002b. Hypermethylation-associated Inactivation of the Cellular Retinol-Binding-Protein 1 Gene in Human Cancer. *Cancer Res* 62: 5902-5.
Evans, T. C., J. Benner, and M. Q. Xu. 1999. The cyclization and polymerization of bacterially expressed proteins using modified self-splicing inteins. *J. Biol. Chem.* 274: 18359-63.
Ferreira, H., D. Sherratt, and L. Arciszewska. 2001. Switching catalytic activity in the XerCD site-specific recombination machine. *J Mol Biol* 312: 45-57.
Fitzmaurice, W. P., A. S. Waldman, R. C. Benjamin, P. C. Huang, and J. J. Scocca. 1984. Nucleotide sequence and properties of the cohesive DNA termini from bacteriophage HP1c1 of Haemophilus influenzae Rd. *Gene* 31: 197-203.
Flashman, S. M. 1978. Mutational analysis of the operators of bacteriophage lambda. *Mol Gen Genet* 166: 61-73.
Folk, W. R. 1981. Phage T4 polynucleotide kinase. *Gene Amplif Anal* 2: 299-311.
Fraga, M. F., and M. Esteller. 2002. DNA methylation: a profile of methods and applications. *Biotechniques* 33: 632, 634, 636-49.
Gentry-Weeks, C., P. S. Coburn, and M. S. Gilmore. 2002. Phages and other mobile virulence elements in gram-positive pathogens. *Curr Top Microbiol Immunol* 264: 79-94.
Georgiou, M., C. P. Georgopoulos, and H. Eisen. 1979. An analysis of Tro phenotype of bacteriophage □. *Virology* 94: 38.
Gildemeister, E. 1921. Ueber das d'Herellesche Phånomen. *Berlin klin. Wochschr.* 58: 1355.
Goldberg, A. L., R. P. Moerschell, C. H. Chung, and M. R. Maurizi. 1994. ATP-dependent protease La (lon) from Escherichia coli. *Methods Enzymol* 244: 350-75.
Goodman, S. D., S. C. Nicholson, and H. A. Nash. 1992. Deformation of DNA during site-specific recombination of bacteriophage lambda: replacement of IHF protein by HU protein or sequence-directed bends. *Proc Natl Acad Sci U S A* 89: 11910-4.
Gopaul, D. N., and G. D. Duyne. 1999. Structure and mechanism in site-specific recombination. *Curr Opin Struct Biol* 9: 14-20.
Gopaul, D. N., F. Guo, and G. D. Van Duyne. 1998. Structure of the Holliday junction intermediate in Cre-loxP site-specific recombination. *Embo J* 17: 4175-87.
Grainge, I., and M. Jayaram. 1999. The integrase family of recombinases: organization and function of the active site. *Mol Microbiol* 33: 449-456.
Grigoriev, P. S., and M. B. Lobocka. 2001. Determinants of segregational stability of the linear plasmid-prophage N15 of Escherichia coli. *Mol Microbiol* 42: 355-68.
Guhathakurta, A., I. Viney, and D. Summers. 1996. Accessory proteins impose site selectivity during ColE1 dimer resolution. *Mol Microbiol* 20: 613-20.
Guo, F., D. N. Gopaul, and G. D. van Duyne. 1997. Structure of Cre recombinase complexed with DNA in a site-specific recombination synapse. *Nature* 389: 40-6.
Guo, F., D. N. Gopaul, and G. D. Van Duyne. 1999. Asymmetric DNA bending in the Cre-loxP site-specific recombination synapse. *Proc Natl Acad Sci U S A* 96: 7143-8.

Hallet, B., L. K. Arciszewska, and D. J. Sherratt. 1999. Reciprocal control of catalysis by the tyrosine recombinases XerC and XerD: an enzymatic switch in site-specific recombination. *Mol Cell* 4: 949-59.
Han, Y. W., R. I. Gumport, and J. F. Gardner. 1993. Complementation of bacteriophage lambda integrase mutants: evidence for an intersubunit active site. *Embo J* 12: 4577-84.
—. 1994a. Mapping the functional domains of bacteriophage lambda integrase protein. *J Mol Biol* 235: 908-925.
—. 1994b. Mapping the functional domains of bacteriophage lambda integrase protein. *J Mol Biol* 235: 908-25.
Hatfull, G. F., and G. J. Sarkis. 1993. DNA sequence, structure and gene expression of mycobacteriophage L5: a phage system for mycobacterial genetics. *Mol Microbiol* 7: 395-405.
Hendrix, R. W. 1988. Tail length determination in double-stranded DNA bacteriophages. *Curr Top Microbiol Immunol* 136: 21-9.
Hendrix, R. W., M. C. Smith, R. N. Burns, M. E. Ford, and G. F. Hatfull. 1999. Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. *Proc Natl Acad Sci U S A* 96: 2192-7.
Herriot, R. M. 1951. Nucleic acid-free T2 virus "ghosts" with specific biological action. *J Bacteriol.* 61: 752.
Hershey, A. D. 1946. Mutation of bacteriophage with respect to type of plaque. *Genetics* 31: 620.
Hershey, A. D., and M. Chase. 1951. Genetic Recombination and heterozygosis in bacteriophage. *Cold Spring Harbor Symp. Quant. Biol.* 16: 471.
—. 1952. Independent functions of viral protein and nucleic acid in growth of bacteriophage. *J. Gen. Physiol.* 36: 39-56.
Hershey, A. D., and R. Rotman. 1949. Gentic recombination between host range and plaque-type mutants of bactreiophage in single bacterial cells. *Genetics* 34: 44.
Hickman, A. B., S. Waninger, J. J. Scocca, and F. Dyda. 1997. Molecular organization in site-specific recombination: the catalytic domain of bacteriophage HP1 integrase at 2.7 A resolution. *Cell* 89: 227-37.
Hodgman, T. C., H. Griffiths, and D. K. Summers. 1998. Nucleoprotein architecture and ColE1 dimer resolution: a hypothesis. *Mol Microbiol* 29: 545-58.
Hoess, R., K. Abremski, S. Irwin, M. Kendall, and A. Mack. 1990. DNA specificity of the Cre recombinase resides in the 25 kDa carboxyl domain of the protein. *J Mol Biol* 216: 873-882.
Hong, J., G. R. Smith, and B. N. Ames. 1971. Adenosine 3'-5'-cyclic monophosphate concentration in the bacterial host regulates the viral decision between lysogeny and lysis. *Proc Natl Acad Sci U S A* 68: 2258.
Hoyt, M. A., D. M. Knight, A. Das, H. I. Miller, and H. Echols. 1982. Control of phage development by stability and synthesis of cII protein: Role of viral cIII and host *hfl*A, *him*A, and *him*D. *Cell* 31: 565.
Huber, K. E., and M. K. Waldor. 2002. Filamentous phage integration requires the host recombinases XerC and Xer D. *Nature* 417: 656-659.
Hutter, B., and T. Dick. 1999. Molecular genetic characterisation of whiB3, a mycobacterial homologue of a Streptomyces sporulation factor. *Res Microbiol* 150: 295-301.
Jacob, F. 1955. Transduction of lysogeny in *Escherichia coli*. *Virology* 1: 207.

Jacob, F., and A. Campbell. 1959. Sur le systéme de répression assurant l'immunité chez les bactéries lysogénes. *Compt. rend. Acad. Sci.* 248: 3219.
Jessop, L., T. Bankhead, D. Wong, and A. M. Segall. 2000. The amino terminus of bacteriophage lambda integrase is involved in protein-protein interactions during recombination. *J Bacteriol* 182: 1024-34.
Johnson, A. D., A. R. Poteete, G. Lauer, R. T. Sauer, G. K. Ackers, and M. Ptashne. 1981. lambda Repressor and cro--components of an efficient molecular switch. *Nature* 294: 217-23.
Johnson, S. K., S. Bhattacharyya, and M. A. Griep. 2000. DnaB helicase stimulates primer synthesis activity on short oligonucleotide templates. *Biochemistry* 39: 736-44.
Kaiser, A. D., and F. Jacob. 1957. Recombination between related temperate bacteriophages and genetic control of immunity and prophage localization. *Virology* 4: 509.
Kim, S., and A. Landy. 1992. Lambda Int protein bridges between higher order complexes at two distant chromosomal loci attL and attR. *Science* 256: 198-203.
Kornberg, A., S. B. Zimmerman, S. R. Kornberg, and J. Josse. 1959. Enzymatic synthesis of deoxyribonucleic acid. VI. Influence of bacteriophage T2 on the synthetic pathway of the host cells. *Proc Natl Acad Sci U S A* 45: 772.
Kourilsky, P. 1973. Lysogenation by bacteriophage lambda: I. Multiple infections and the lysogenation response. *Mol. Gen. Genet.* 122: 183.
Krueger, A. P. 1931. The sorption of bacteriophage by living and dead susceptible bacteria. *J. Gen. Pphysiol.* 14: 493.
Kuo, T. T., Y. S. Chao, Y. H. Lin, B. Y. Lin, L. F. Liu, and T. Y. Feng. 1987. Integration of the DNA of filamentous bacteriophage Cf1t into chromosomal DNA of its host. *J Virol* 61: 60-65.
Kwon, H. J., R. Tirumalai, A. Landy, and T. Ellenberger. 1997. Flexibility in DNA recombination: structure of the lambda integrase catalytic core. *Science* 276: 126-31.
Landy, A., R. H. Hoess, K. Bidwell, and W. Ross. 1979. Site-specific recombination in bacteriophage lambda: structural features of recombining sites. *Cold Spring Harb Symp Quant Biol* 43: 1089-97.
Landy, A., P. L. Hsu, W. Ross, and M. Buraczynska. 1980. Site-specific recombination in bacteriophage lambda: structural analyses of reactive DNA sequences. *Am J Trop Med Hyg* 29: 1099-106.
Laskowska, E., D. Kuczynska-Wisnik, J. Skorko-Glonek, and A. Taylor. 1996. Degradation by proteases Lon, Clp and HtrA, of Escherichia coli proteins aggregated in vivo by heat shock; HtrA protease action in vivo and in vitro. *Mol Microbiol* 22: 555-71.
Lederberg, E. M., and J. Lederberg. 1953. Genetic studies of lysogenicity in *Escherichia coli*. *Genetics* 38: 51.
Lee, J., M. Jayaram, and I. Grainge. 1999. Wild-type Flp recombinase cleaves DNA in trans. *Embo J* 18: 784-791.
Lee, M. H., and G. F. Hatfull. 1993. Mycobacteriophage L5 integrase-mediated site-specific integration in vitro. *J Bacteriol* 175: 6836-41.
Lee, M. H., L. Pascopella, W. R. Jacobs, Jr., and G. F. Hatfull. 1991. Site-specific integration of mycobacteriophage L5: integration- proficient vectors for Mycobacterium smegmatis, Mycobacterium tuberculosis, and bacille Calmette-Guerin. *Proc Natl Acad Sci U S A* 88: 3111-5.

Lehnherr, H., and J. Meyer. 2000. Enterobacteria phage P1 (Myoviridae). Pages 455-461 in A. Granoff and R. G. Webster, eds. *Encyclopedia of Virology*. Academic Press, London.
Levinthal, C. 1954. Recombination in phage T2; its relation to heterozygosis and growth. *Proc Natl Acad Sci U S A* 42: 394.
Lewis, J. A., and G. F. Hatfull. 2001. Control of directionality in integrase-mediated recombination: examination of recombination directionality factors (RDFs) including Xis and Cox proteins. *Nucleic Acids Res* 29: 2205-16.
Lewis, M., A. Jeffrey, J. Wang, R. Ladner, M. Ptashne, and C. O. Pabo. 1983. Structure of the operator-binding domain of bacteriophage lambda repressor: implications for DNA recognition and gene regulation. *Cold Spring Harb Symp Quant Biol* 47: 435-40.
Lieb, M. 1953. The establishment of lysogenicity in *Escherichia coli*. *J Bact* 65: 642.
Lindqvist, B. H., and E. W. Six. 1971. Replication of bacteriophage P4 DNA in nonlysogenic host. *Virology* 43: 1-7.
Liu, T., and E. Haggard-Ljungquist. 1999. The transcriptional switch of bacteriophage WPhi, a P2-related but heteroimmune coliphage. *J Virol* 73: 9816-26.
Luetke, K. H., and P. D. Sadowski. 1998. Determinants of the position of a Flp-induced DNA bend. *Nucleic Acids Res* 26: 1401-1407.
Lwoff, A., and A. Gutmann. 1950. Recherches sur un Bacillus megathérium lysogéne. *Ann. Inst. Pasteur* 78: 711.
Lwoff, A., L. Siminovitch, and N. Kjeldgaard. 1950. Induction de la production de bactériophages chez une bactérie lysogéne. *Ann. Inst. Pasteur* 79: 815.
Maniatis, T., and M. Ptashne. 1973. Multiple repressor binding at the operators in bacteriophage lambda. *Proc Natl Acad Sci U S A* 70: 1531-5.
Mann, B. A., and J. M. Slauch. 1997. Transduction of low-copy number plasmids by bacteriophage P22. *Genetics* 146: 447-56.
Meselson, M., and F. W. Stahl. 1958. The replication of DNA in *Escherichia coli*. *Proc Natl Acad Sci U S A* 44: 671.
Meselson, M., F. W. Stahl, and J. Vinograd. 1957. Equilibrium sedimentation of macromolecules in density gradients. *Proc Natl Acad Sci U S A* 43: 581.
Meselson, M., and J. J. Weigle. 1961. Chromosome breakage accompanying genetic recombination in bacteriophage. *Proc Natl Acad Sci U S A* 47: 857.
Mhammedi-Alaoui, A., M. Pato, M. J. Gama, and A. Toussaint. 1994. A new component of bacteriophage Mu replicative transposition machinery: the Escherichia coli ClpX protein. *Mol Microbiol* 11: 1109-16.
Miao, E. A., and S. I. Miller. 1999. Bacteriophages in the evolution of pathogen-host interactions. *Proc Natl Acad Sci U S A* 96: 9452-4.
Miller, G., and M. Feiss. 1988. The bacteriophage lambda cohesive end site: isolation of spacing/substitution mutations that result in dependence on Escherichia coli integration host factor. *Mol Gen Genet* 212: 157-65.
Minter, S. J., G. M. Clore, A. M. Gronenborn, and R. W. Davies. 1986. Cooperative DNA binding by lambda integration protein--a key component of specificity. *Eur J Biochem* 161: 727-31.
Mizuuchi, M., and K. Mizuuchi. 1993. Target site selection in transposition of phage Mu. *Cold Spring Harbor Symp. Quant. Biol.* 58: 515-523.
Molle, V., W. J. Palframan, K. C. Findlay, and M. J. Buttner. 2000. WhiD and WhiB, homologous proteins required for different stages of sporulation in Streptomyces coelicolor A3(2). *J Bacteriol* 182: 1286-95.

Mozola, M. A., D. L. Carver, and D. L. Friedman. 1985. Unproductive infection of Phi80. *Virology* 140: 328-41.
Mulder, N. J., H. Zappe, and L. M. Steyn. 1999. Characterization of a Mycobacterium tuberculosis homologue of the Streptomyces coelicolor whiB gene. *Tuber Lung Dis* 79: 299-308.
Nash, H. A. 1975. Integrative recombination in bacteriophage lambda: analysis of recombinant DNA. *J Mol Biol* 91: 501-14.
—. 1977. Integration and excision of bacteriophage lambda. *Curr Top Microbiol Immunol* 78: 171-99.
—. 1981. Integration and excision of bacteriophage lambda: the mechanism of conservation site specific recombination. *Annu Rev Genet* 15: 143-67.
—. 1990. Bending and supercoiling of DNA at the attachment site of bacteriophage lambda. *Trends Biochem Sci* 15: 222-7.
Nash, H. A., and C. A. Robertson. 1981. Purification and properties of the Escherichia coli protein factor required for lambda integrative recombination. *J Biol Chem* 256: 9246-53.
Nesbit, C. E., M. E. Levin, M. K. Donnelly-Wu, and G. F. Hatfull. 1995. Transcriptional regulation of repressor synthesis in mycobacteriophage L5. *Mol Microbiol* 17: 1045-56.
Noble, J. A., M. A. Innis, E. V. Koonin, K. E. Rudd, F. Banuett, and I. Herskowitz. 1993. The Escherichia coli hflA locus encodes a putative GTP-binding protein and two membrane proteins, one of which contains a protease-like domain. *Proc Natl Acad Sci U S A* 90: 10866-70.
Numrych, T. E., R. I. Gumport, and J. F. Gardner. 1992. Characterization of the bacteriophage lambda excisionase (Xis) protein: the C-terminus is required for Xis-integrase cooperativity but not for DNA binding. *Embo J* 11: 3797-806.
Nunes-Duby, S. E., M. A. Azaro, and A. Landy. 1995. Swapping DNA strands and sensing homology without branch migration in lambda site-specific recombination. *Curr Biol* 5: 139-48.
Nunes-Duby, S. E., H. J. Kwon, R. S. Tirumalai, T. Ellenberger, and A. Landy. 1998. Similarities and differences among 105 members of the Int family of site-specific recombinases. *Nucleic Acids Res* 26: 391-406.
Oyaski, M., and G. F. Hatfull. 1992. The cohesive ends of mycobacteriophage L5 DNA. *Nucleic Acids Res* 20: 3251.
Pan, H. D., D. Clary, and P. D. Sadowski. 1991. Identification of the DNA binding domain of Flp recombinase. *J Biol Chem* 266: 11347-11354.
Park, K., S. Mukhopadhyay, and D. K. Chattoraj. 1998. Requirements for and regulation of origin opening of plasmid P1. *J Biol Chem* 273: 24906-11.
Patsey, R. L., and M. F. Bruist. 1995. Characterization of the interaction between the lambda intasome and attB. *J Mol Biol* 252: 47-58.
Pedulla, M. L., and G. F. Hatfull. 1998. Characterization of the mIHF gene of Mycobacterium smegmatis. *J Bacteriol* 180: 5473-7.
Pedulla, M. L., M. H. Lee, D. C. Lever, and G. F. Hatfull. 1996. A novel host factor for integration of mycobacteriophage L5. *Proc Natl Acad Sci U S A* 93: 15411-6.
Peeters, B. P., J. G. Schoenmakers, and R. N. Konings. 1987. Comparison of the DNA sequences involved in replication and packaging of the filamentous phages IKe and Ff (M13, fd, and f1). *DNA* 6: 139-47.
Pena, C. E., J. M. Kahlenberg, and G. F. Hatfull. 2000. Assembly and activation of site-specific recombination complexes. *Proc Natl Acad Sci U S A* 97: 7760-5.

Peña, C. E., M. H. Lee, M. L. Pedulla, and G. F. Hatfull. 1997. Characterization of the mycobacteriophage L5 attachment site, attP. *J Mol Biol* 266: 76-92.
Peña, C. E., J. Stoner, and G. F. Hatfull. 1998. Mycobacteriophage D29 integrase-mediated recombination: specificity of mycobacteriophage integration. *Gene* 225: 143-51.
Peña, C. E., J. E. Stoner, and G. F. Hatfull. 1996. Positions of strand exchange in mycobacteriophage L5 integration and characterization of the attB site. *J Bacteriol* 178: 5533-6.
Peña, C. E. A., and G. F. Hatfull. 2000. Structure and assembly or recombinagenic complexes. *Manuscript submitted*.
Peña, C. E. A., J. M. Kahlenberg, and G. F. Hatfull. 1999. Protein-DNA complexes in mycobacteriophage L5 integrative recombination. *J Bacteriol* 181: 454-61.
Perler, F. B. 1998. Protein splicing of inteins and hedgehog autoproteolysis: Structure, function, and evolution. *Cell* 92: 1-4.
Phizicky, E. M., and J. W. Roberts. 1980. Kinetics of RecA protein-directed inactivation of repressors of phage lambda and phage P22. *J Mol Biol* 139: 319-28.
Ptashne, M. 1973. Repressor, operators, and promoters in bacteriophage lambda. *Harvey Lect* 74: 143-71.
—. 1987. *A Genetic Switch*. Blackwel Science Ltd and Cell Press, Oxford/Cambridge.
Ptashne, M., K. Backman, M. Z. Humayun, A. Jeffrey, R. Maurer, B. Meyer, and R. T. Sauer. 1976. Autoregulation and function of a repressor in bacteriophage lambda. *Science* 194: 156-61.
Ptashne, M., and N. Hopkins. 1968. The operators controlled by the lambda phage repressor. *Proc Natl Acad Sci U S A* 60: 1282-7.
Ptashne, M., A. D. Johnson, and C. O. Pabo. 1982. A genetic switch in a bacterial virus. *Sci Am* 247: 128-30, 132, 134-40.
Ravin, V., N. Ravin, S. Casjens, M. E. Ford, G. F. Hatfull, and R. W. Hendrix. 2000. Genomic sequence and analysis of the atypical temperate bacteriophage N15. *J Mol Biol* 299: 53-73.
Ravin, V. K., and M. G. Shulga. 1970. The evidence for extrachromosomal location of prophage N15. *Virology* 40: 800-807.
Recchia, G. D., M. Aroyo, D. Wolf, G. Blakely, and D. J. Sherratt. 1999. FtsK-dependent and -independent pathways of Xer site-specific recombination. *Embo J* 18: 5724-34.
Recchia, G. D., and D. J. Sherratt. 1999. Conservation of xer site-specific recombination genes in bacteria. *Mol Microbiol* 34: 1146-8.
Redinbo, M. R., L. Stewart, P. Kuhn, J. J. Chapoux, and W. G. Hol. 1998. Crystal structures of human topoisomerase I in covalent and non-covalent complexes with DNA. *Science* 279: 1504-1513.
Rice, P. A. 2000. Structure of Flp-DNA comlpex. *Mol Cell* 6: 885-897.
Rice, P. A., S. W. Yang, K. Mizuuchi, and H. Nash. 1996. Crystal structure of an IHF-DNA complex: a protein-induced DNA u-turn. *Cell* 87: 1295-1306.
Richet, E., P. Abcarian, and H. A. Nash. 1988. Synapsis of attachment sites during lambda integrative recombination involves capture of a naked DNA by a protein-DNA complex. *Cell* 52: 9-17.
Roberts, J. W., and C. W. Roberts. 1975. Proteolytic cleavage of bacteriophage lambda repressor in induction. *Proc Natl Acad Sci U S A* 72: 147-51.
Roberts, J. W., C. W. Roberts, and N. L. Craig. 1978. Escherichia coli recA gene product inactivates phage lambda repressor. *Proc Natl Acad Sci U S A* 75: 4714-8.

Sarkar, D., M. Radman-Livaja, and A. Landy. 2001. The small DNA binding domain of lambda integrase is a context-sensitive modulator of recombinase function. *Embo J* 20: 1203-1212.
Schlesinger, M. 1932a. Ueber die Bindung des Bakteriophagen an homologe Bakterien I. Die Unterscheidung von Gruppen von verschiedener Binungsaffinioder Irreversibilitaet der Bindung. *Z. Hyg. Infekttionskrankh* 114: 136-148.
—. 1932b. Ueber die Bindung des Bakteriophagen an homologe Bakterien II. Quantitative Untersuchungen ueber die Bindungsgeschwindigkiet und die Saettigung. Berechnung der Teilchengroesse des Bakteriophagen aus deren Mengen. *Z. Hyg. Infekttionskrankh* 114: 149-160.
—. 1934. Zur Frage der chemischen Zusammensetzung des Bakteriophagen. *Biochem. Z.* 273: 306.
Schwartz, C. J., and P. D. Sadowski. 1990. Flp protein of 2 mu circle plasmid of yeast induces multiple bends in Flp recognition target site. *J Mol Biol* 216: 289-298.
Schweder, T., K. H. Lee, O. Lomovskaya, and A. Matin. 1996. Regulation of Escherichia coli starvation sigma factor (sigma s) by ClpXP protease. *J Bacteriol* 178: 470-6.
Segall, A. M., S. G. Goodman, and H. Nash. 1994. Architechural elements in nucleoprotein complexes: interchangeability of specific and non-specific DNA binding proteins. *Embo J* 13: 4536-4548.
Segall, A. M., and H. A. Nash. 1993. Synaptic intermediates in bacteriophage lambda site-specific recombination: integrase can align pairs of attachment sites. *Embo J* 12: 4567-76.
Sherrat, D. J., and D. B. Wigley. 1998. Conserved themes but novel activities in recombinases and topoisomerases. *Cell* 93: 149-152.
Sherratt, D. J., I. F. Lau, and F. X. Barre. 2001. Chromosome segregation. *Curr Opin Microbiol* 4: 653-9.
Shimatake, H., and M. Rosenberg. 1981. Purified l regulatory protein cII positively activates promoters for lysogenic development. *Nature* 272: 128.
Six, E. W. 1963. A defective phage depending on phage P2. *Bacteriol. Proc.* 138.
Steiner, W., G. Liu, W. D. Donachie, and P. Kuempel. 1999. The cytoplasmic domain of FtsK protein is required for resolution of chromosome dimers. *Mol Microbiol* 31: 579-83.
Stewart, L., M. R. Redinbo, X. Qiu, W. G. Hol, and J. J. Champoux. 1998. A model for the mechanism of human topoisomerase I. *Science* 279: 1534-41.
Streisinger, G., R. H. Edgar, and G. H. Denhardt. 1964. The gross structure of the genome of phage T4. I. The circularity of the linkage map. *Proc Natl Acad Sci U S A*: 775.
Subramanya, H. S., L. K. Arciszewska, R. A. Baker, L. E. Bird, D. J. Sherratt, and D. B. Wigley. 1997. Crystal structure of the site-specific recombinase, XerD. *Embo J* 16: 5178-87.
Summers, D. 1998. Timing, self-control and a sense of direction are the secrets of multicopy plasmid stability. *Mol Microbiol* 29: 1137-45.
Svarchevsky, A. N., and V. N. Rybchin. 1984. Physical mapping of plasmid N15 DNA. *Mol Genet (Moscow)* N10: 16-22.
Symonds, N., A. Tousant, P. Van de Putte, and M. Howe. 1987. *Phage Mu*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
Thompson, J. F., L. M. de Vargas, S. E. Skinner, and A. Landy. 1987a. Protein-protein interactions in a higher-order structure direct lambda site-specific recombination. *J Mol Biol* 195: 481-93.

Thompson, J. F., and A. Landy. 1988. Empirical estimation of protein-induced DNA bending angles: applications to lambda site-specific recombination complexes. *Nucleic Acids Res* 16: 9687-705.
Thompson, J. F., L. Moitoso de Vargas, C. Koch, R. Kahmann, and A. Landy. 1987b. Cellular factors couple recombination with growth phase: characterization of a new component in the lambda site-specific recombination pathway. *Cell* 50: 901-8.
Thorpe, H. M., and M. C. Smith. 1998. In vitro site-specific integration of bacteriophage DNA catalyzed by a recombinase of the resolvase/invertase family. *Proc Natl Acad Sci U S A* 95: 5505-10.
Thorpe, H. M., S. E. Wilson, and M. C. Smith. 2000. Control of directionality in the site-specific recombination system of the streptomyces phage phiC31 [In Process Citation]. *Mol Microbiol* 38: 232-41.
Tirumalai, R. S., E. Healey, and A. Landy. 1997. The catalytic domain of lambda site-specific recombinase. *Proc Natl Acad Sci U S A* 94: 6104-9.
Tirumalai, R. S., H. J. Kwon, E. H. Cardente, T. Ellenberger, and A. Landy. 1998. Recognition of core-type DNA sites by lambda integrase. *J Mol Biol* 279: 513-27.
Twort, F. W. 1915. An Investigation on the Nature of Ultramicroscopic Viruses. *Lancet* II.
Unger, R., and A. Clark. 1972. Interaction of the recombination pathways of bacteriophage lambda and its host Escherichia coli K12: Effects on exonuclease V activity. *Mol. Gen. Genet.* 187.
Van Duyne, G. D. 2001. A structural view of cre-loxp site-specific recombination. *Annu Rev Biophys Biomol Struct* 30: 87-104.
—. 2002. A structural view of tyrosine recombinase site-specific recombination. Pages 93-117 in N. L. Craig, R. Craiige, M. Gellert, and A. M. Lambowitz, eds. *Mobile DNA II*. ASM Press, Washington, D.C.
Weinfeld, M., M. Liuzzi, and M. C. Paterson. 1989. Enzymatic analysis of isomeric trithymidylates containing ultraviolet light-induced cyclobutane pyrimidine dimers. II. Phosphorylation by phage T4 polynucleotide kinase. *J Biol Chem* 264: 6364-70.
Weisberg, R. A., L. W. Enquist, C. Foeller, and A. Landy. 1983. Role for DNA homology in site-specific recombination. The isolation and characterization of a site affinity mutant of coliphage lambda. *J Mol Biol* 170: 319-42.
Wojciak, J. M., D. Sarkar, A. Landy, and R. T. Clubb. 2002. Arm-site binding by lambda -integrase: solution structure and functional characterization of its amino-terminal domain. *Proc Natl Acad Sci U S A* 99: 3434-9.
Wollman, E. L. 1953. Sur le determinisme génétique de la lysogenie. *Ann. Inst. Pasteur* 84: 281.
Wollman, E. L., and F. Jacob. 1954. Lysogénie et recombinaison génétique chez *Escherichia coli* K12. *Compt. rend.* 239: 455.
—. 1957. Sur les processus de conjugation et de récombinasion chez Escherichia coli. II. La localisation chromosomoqie du prophage et les conséquences de l'induction zygotique. *Ann. Inst. Pasteur* 93: 323.
Wu, Z., R. I. Gumport, and J. F. Gardner. 1998. Defining the structural and functional roles of the carboxyl region of the bacteriophage lambda excisionase (Xis) protein. *J Mol Biol* 281: 651-61.

Xiong, G., P. Oepen, R. Geiben, A. H. el-Idrissi, and F. Lutz. 1996. Plasmids containing cos ends inhibit the replication of phage phi CTX in Pseudomonas aeruginosa. *Virus Res* 41: 77-87.
Xu, M., and T. C. Evans. 2001. Intien-mediated ligation and cyclization of expressed proteins. *Methods* 24: 257-277.
Xu, S. Y., and M. Feiss. 1991. Structure of the bacteriophage lambda cohesive end site. Genetic analysis of the site (cosN) at which nicks are introduced by terminase. *J Mol Biol* 220: 281-92.
Yamamoto, K., B. Low, S. A. Rutherford, M. Rajagopalan, and M. V. V. S. Madiraju. 2001. The *Mycobacterium avium-intracellulare* complex dnaB locus and protein intein splicing. *Biochem. and Biophys. Res. Comm.* 280: 898-903.
Yin, S., W. Bushman, and A. Landy. 1985. Interaction of the lambda site-specific recombination protein Xis with attachment site DNA. *Proc Natl Acad Sci U S A* 82: 1040-4.
Yu, X., M. J. Jezewska, W. Bujalowski, and E. H. Egelman. 1996. The hexameric E. coli DnaB helicase can exist in different Quaternary states. *J Mol Biol* 259: 7-14.